UNIVERSIDAD POLITÉCNICA DE MADRID

ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INFORMÁTICOS

MODELO DE FUSIÓN HOLÓNICA DE INFORMACIÓN PARA LA MEJORA DE LA CALIDAD DE LA INFORMACIÓN EN REDES PEER-TO-PEER CON RECURSOS LIMITADOS

TESIS PRESENTADA PARA OPTAR AL TÍTULO DE DOCTOR

Horacio Paggi Straneo

Ing. en Computación, Mag. en Ing. en Computación, Mag. en Ing. Matemática,
Universidad de la República, Uruguay

Directores:

Dr. Francisco Javier Soriano Camino y Dr. Juan Alfonso Lara Torralbo

Madrid, 2018

Agradecimientos

Son muchas las personas e instituciones a las que tengo que estar agradecido por el apoyo y la ayuda que me dieron en distintos momentos mi doctorado. Perdón si me olvido de algún nombre, ha pasado mucho tiempo desde que comencé esta aventura, pero los gestos y aliento no los olvido.

Primero que nada, a mis tutores actuales Dr. Francisco Javier Soriano Camino y Dr. Juan Alfonso Lara Torralbo, gracias a cuyo acompañamiento y apoyo he logrado llegar hasta este momento, algo que en cierto momento me pareció inalcanzable. También quisiera agradecer a quienes lo fueron, Dr. José Crespo Del Arco y, póstumamente, al Dr. Fernando Alonso Amo, que deseaba fervientemente estar presente en el momento de la lectura de tesis y a quien la vida le dio una mala sorpresa y ya no pudo hacerlo.

Segundo, a todo el personal del Cettico (especialmente a Agustín Rodríguez Rodríguez por su ayuda en los aspectos administrativos y a Rachel Elliott por el esfuerzo hecho en llegar con las traducciones a tiempo) y en general al de la UPM-ETSIInf., en especial a Alicia Andrés por su ayuda en asuntos administrativos: gracias por haber resuelto los problemas que se generan al trabajar a tantos miles de kilómetros.

Tercero, a mis compañeros de la Universidad ORT, en especial a Carmen Signorelli y Daniel Pereira, que tanto me ayudaron en los trámites y vueltas necesarios sobre todo al comienzo del doctorado, y a los doctores Gerardo Matturro y Pedro Salvetto por sus invaluable buenos consejos respecto al desarrollo del doctorado.

Cuarto, a mis primos, Pablo Montans Leites y Carlos Montans Leites, que fueron apoyo de mi casa cada vez que tuve que ausentarme, y a Horacio Ibarra y familia, que supieron acompañarme también en momentos muy duros.

Quinto y último, pero de capital importancia, al Programa de Desarrollo Tecnológico (PDT) del Ministerio de Educación y Cultura, Uruguay, sin cuyo financiamiento este trabajo no hubiese sido posible, y en especial a José Carrera y Graciela Burgeño que me acompañaron y aconsejaron en la gestión de dicha beca-préstamo.

Montevideo, 28 de noviembre de 2018.

A mi madre

Resumen

Las redes de pares (*peer-to-peer*) formadas por elementos heterogéneos a menudo dependen de recursos que son limitados (en términos de energía, ancho de banda, tiempo de comunicación, etc.). Debido a esto, y con el objetivo de cuidar esos recursos, los componentes de la red tienen que comportarse necesariamente en forma inteligente para hacer un uso eficiente, si no óptimo, de los mismos. Además, dichos componentes deben ser capaces de coordinar sus acciones con el objetivo de alcanzar sus metas en presencia de datos que pueden ser vagos (por ejemplo, por estar expresados con expresiones ambiguas o imprecisas, por ejemplo “temperatura alta”) o inciertos (datos con cierto error: como en el caso de error de medición), respetando su autonomía. En la investigación en el campo de las redes *peer-to-peer* poco se han estudiado los efectos de la mejora de la calidad de la información en el funcionamiento de la red, en especial si los recursos son limitados. En este trabajo se describe un modelo de sistema multi-agente que realiza fusión de la información y que se basa en pares (*peers*) colaborativos para mejorar la calidad de la información manejada por los agentes. El modelo se basa en una estrategia consistente en: a) consultar a los *peers* que han tenido históricamente mejor comportamiento para un agente consultante y un tipo de información dados; y b) si la calidad de la información a responder es menor que la del consultante, contestar sólo algunas veces, con una probabilidad que va decreciendo en el tiempo, en un efecto similar al del recocido simulado donde el sistema puede pasar a un estado de mayor energía con una probabilidad no nula que va decreciendo. El modelo propuesto se apoya además en una métrica definida *ad hoc* para la calidad de la información. Dicha métrica incorpora varias dimensiones que generalmente son ignoradas en la literatura, tales como la cantidad de respuestas recibidas a una consulta, la vaguedad de la información devuelta, la cantidad de *peers* consultados, etc. Se demuestra que el sistema resultante posee características de auto-organización que permiten que se formen estructuras holónicas de manera efectiva (esto es, que la probabilidad de que no se formen dichas estructuras tiende a cero al crecer el número de agentes). Se presentan además tres casos de estudio: uno asociado a la tipificación de sitios web como de *phishing* a partir de sus características, otro relativo a la determinación del nivel diario de ozono presente en la atmósfera y el tercero sobre la se-

lección de un automóvil en base a sus características (precio, confort, etc.). Los tres casos de estudio, dadas sus características, son de alguna forma complementarios. Se reportan los experimentos realizados sobre una implementación simulada del modelo, habiéndose realizado las pruebas suficientes como para que las conclusiones sean estadísticamente válidas. Los resultados muestran que el modelo propuesto permite tener mejor rendimiento (medido como cantidad de *time-outs*, cantidad de mensajes necesarios para obtener una respuesta de cierta calidad, etc.) que si se usara una red *peer-to-peer* plana (es decir, sin la inteligencia propia del modelo propuesto) cuando los recursos (mensajes, energía, etc.) son limitados. Finalmente, se presenta un ejemplo en el que se muestran paso a paso las estructuras holónicas que se forman al aplicar el modelo propuesto a la toma de decisiones asociada a la compra de un vehículo. Un sistema de este tipo tiene un amplio espectro de dominios de aplicación, desde sistemas de recomendación móviles a aplicaciones para la toma de decisiones donde se busca alcanzar una meta común en entornos críticos con escasez de recursos o *peers* que pueden dejar de contestar en cualquier momento.

Palabras Clave: Modelo de calidad de la información, sistemas P2P adaptativos, formación de los holones, redes con recursos limitados, manejo de información incierta, fusión holónica de información, *soft sensors*.

Abstract

Peer-based networks (peer-to-peer) formed by heterogeneous elements often depend on limited resources (power, communication bands, communications time, etc.). Due to this the networks components have to behave in an intelligent way in order to use efficiently these resources. Also the aforementioned components have to be able of to coordinate their actions to reach their individual goals in the presence of data which can be vague (for example, due to being expressed as ambiguous or imprecise expressions, for example “high temperature”) or uncertain (data with a certain error, as in the case of a measurement error) while respecting their existing autonomies. In the research field of the peer-to-peer networks the effects of the improvement of the information quality in the network working has been scarcely studied, especially when the resources are limited. In this work a model of multi-agent is described which performs information fusion and which is collaborative peers-based to improve the quality of the information processed by the agents. The model is based in a strategy consisting of a) consulting the peers which had the better behavior for a querying agent and an information type (e.g. a certain data field) and b) if the quality of the information to be answered is lesser than the consultant, there is a probability, which decreases through time, that it is answered anyway. This imitates the effect of the simulated annealing where the system can pass to a state of higher energy with a non-null probability which decreases through time. The proposed model is based also in an information quality metric defined ad hoc which incorporates several dimensions that generally are ignored in the literature such as the number of received answers for a given query, the vagueness of the replied information, the number of consulted peers, etc.. It is proofed that the system has auto-organization characteristics which enables the formation of holonic structures and that they are effectively formed (or, equivalently, that the probability that they are not obtained tends to 0 when the number of agents grows). Three study cases are presented: one related to the classification of websites as phising or safe sites using their characteristics, another one associated to the measurement of the daily ozone level in the atmosphere and a third one about the selection of a car based on its properties (price, comfort, etc.). Because of their characteristics the three cases are somehow complementary. The performed experiments used a simulated implementation

of the model. The number of tests was enough as to ensure the statistical validity of the conclusions. The results are reported and are positive as they show that the proposed model allows having better performance (measured as number of time-outs, number of messages needed to get an answer of a minimum, preset quality, etc.) respect to a flat peer-to-peer network (v.g. a network without the intelligence of the proposed model here) when the resources (messages, energy, etc.) are limited. Finally, is presented an example showing step-by-step the formation of holonic structures when applying the model to the decision making – to decide if a vehicle is purchased. A system of this kind has an ample spectrum of application domains ranging from mobile recommending systems to applications for the decision making where a common goal is pursued in critical environments with scarce resources or with peers that can stop answering at any moment, as it happens in many critical environments.

Keywords: information quality model, adaptive P2P systems, holon formation, limited resources networks, uncertain information handling, holonic information fusion, soft sensors.

Siglas y abreviaturas

ANFIS	<i>Adaptive Network-based Fuzzy Inference System</i>
D-S	Dempster-Shafer
DNA	<i>Deoxyribonucleic acid (ácido desoxirribonucleico)</i>
DNS	<i>Domain Name System</i>
DTN	<i>Delay Tolerant Networks</i>
DQ	<i>Data Quality Generalized</i>
GIS	<i>Intelligent Space</i>
GPRS	<i>General Packet Radio Services</i>
GPS	<i>Global Positioning System</i>
IF	<i>Information Fusion</i>
IoT	<i>Internet of Things</i>
IS	<i>Intelligent Space</i>
IQ	<i>Information Quality</i>
JDL	<i>Joint Directors of Laboratories</i>
P2P	<i>Peer-to-peer</i>
PCA	<i>Principal Component Analysis</i>
RBF	<i>Radial Basis Function</i>
RNA	Red Neuronal Artificial
SAW	<i>Situation Awareness</i>
SS	<i>Soft Sensor</i>
SOM	<i>Self-organizing Network</i>
TO	<i>Time-Out</i>
VANets	<i>Vehicular ad hoc networks</i>
WSN	<i>Wireless Sensor Network</i>

1. Introducción, Motivación y Objetivos	1
1.1. Motivación	5
1.2. Objetivos	8
1.3. Aportaciones	10
1.4. Contribuciones	10
1.5. Organización del documento	11
 2. Estado de la cuestión	 13
2.1. Calidad de la información	13
2.2. Indeterminación, vaguedad e incertidumbre	14
2.3. Redes de recursos limitados	17
2.4. Sistemas <i>Peer-to-Peer</i>	18
2.4.1. Generalidades	18
2.4.2. Redes P2P en el contexto de la Tesis	22
2.5. Objetos colaborativos y espacios inteligentes	22
2.6. <i>Sensemaking</i>	24
2.7. <i>Soft sensors</i>	25
2.7.1. <i>Soft sensors</i> distribuidos	26
2.7.2. Sistemas de <i>soft sensors</i> basados en la naturaleza	27
2.8. Fundamentos de las estructuras holónicas	28
2.8.1. Holones	28
2.8.2. Informones	45
2.8.3. Entidades	48
2.9. Fusión de información	49
2.9.1. Definición	49
2.9.2. Generalidades	52
2.9.3. El estándar JDL	54
2.9.4. Fusión de información holónica	56
2.10. Auto-organización	56
2.10.1. Características básicas de la auto-organización en el sistema	58
2.10.2. Características adicionales de redes P2P auto-organizadas	60

3. Planteamiento del problema y preguntas de investigación	61
4. Trabajo relacionado	63
4.1. Uso de la fusión de información para reducir la incertidumbre y vaguedad	63
4.2. Dimensiones de la IQ consideradas	66
4.3. Formas y objetivos de la fusión de datos	69
4.3.1. Clasificación de la IF según la topología	69
4.3.2. Clasificación de la IF según las funciones usadas	76
4.3.3. Clasificación de la IF según los objetivos	78
4.4. Redes con recursos restringidos	79
4.5. Oportunidades de investigación encontradas	80
5. Descripción de la solución	83
5.1. Aspectos generales de la propuesta	84
5.2. Fórmula de calidad usada	88
5.3. Dinámica de un sistema con este modelo	91
5.4. Parámetros del modelo	97
5.5. Descripciones de los procesos y comportamientos	101
6. Trabajo experimental	107
6.1. Descripción general de los experimentos	107
6.2. Medida del rendimiento	109
6.3. Generalidades del análisis estadístico realizado	110
6.4. Caso A: Sensores (<i>Soft sensors</i>)	112
6.4.1. Generalidades	112
6.4.2. Resultados obtenidos	112
6.4.3. Análisis estadístico	115
6.5. Caso B: Automóviles	117
6.5.1. Generalidades	117
6.5.2. Resultados obtenidos	117
6.5.3. Análisis estadístico	123
6.5.4. Discusión de los resultados	129
6.6. Caso C: Sitios web	130
6.6.1. Generalidades	130
6.6.2. Resultados obtenidos	131
6.6.3. Análisis estadístico	133
6.6.4. Discusión de los resultados.	141
6.7. Conclusiones estadísticas generales	142
7. Formación de holones	143
7.1. Justificación teórica	143
7.2. Caracterización de la auto-organización en el sistema	148
7.3. Caracterización para la autonomía	149
7.4. Ejemplo de formación de holones	151
7.4.1. Dinámica del sistema	151
7.4.2. Formación de los holones	161

8. Conclusiones	165
9. Trabajo futuro	169
Bibliografía	172
ANEXOS	191
Anexo A: Datos del caso de estudio sobre niveles de ozono	195
Anexo B: Datos del caso de estudio de compra de vehículo	203
Anexo C: Datos del caso de estudio de clasificación de sitios web	208

ÍNDICE DE FIGURAS

1.	Esquema cabeza-cuerpo del holón	36
2.	Influencia de los informones en los diferentes niveles	47
3.	Dinámica del sistema - Paso 1	93
4.	Dinámica del sistema - Paso 2	93
5.	Dinámica del sistema - Paso 3	94
6.	Dinámica del sistema - Paso 4	94
7.	Diagrama de flujo del comportamiento de un agente del sistema	98
8.	Diagrama de flujo del comportamiento de un agente externo al sistema.	99
9.	Diagrama de flujo del comportamiento del agente Ω	100
10.	Consumo de mensajes en el sistema al aumentar los agentes.	114
11.	TOs al aumentar los agentes.	114
12.	Consumo de menssajes al aumentar los mensajes.	115
13.	TOs del sistema al aumentar los mensajes.	115
14.	Promedio de éxitos estrictos para un número mínimo dado de mensajes.	119
15.	Número promedio de mensajes requeridos para alcanzar una calidad mínima dada (todos los agentes).	120
16.	Número promedio de TOs para una cantidad de mensajes dada.	120
17.	Promedio de éxitos para un número de agentes mínimo.	121
18.	Promedio de TOs para distintas cantidades de agentes.	121
19.	Promedio de mensajes necesarios para obtener una calidad dada para 30 agentes.	122
20.	Frecuencia con que se obtiene una calidad mínima.	122
21.	Frecuencia promedio de una calidad mínima dada (todos los agentes).	131
22.	Promedio de TOs para distintos números de mensajes.	132
23.	Promedio de time-outs para distintos números de agentes.	132
24.	Número promedio de éxitos estrictos para una cantidad máxima de agentes.	133
25.	Promedio de éxitos estrictos	133
26.	Caso de comunicación anormal	148
27.	Diagrama de la dinámica del ejemplo	159
28.	Cronología de los eventos del ejemplo	160
29.	Evolución de los holones	161
30.	Otra posible evolución de los holones	162

31.	Vista parcial de las relaciones de dependencia del caso A.	201
32.	Composición de los campos caso B.	207
33.	Composición de los campos caso C.	212

ÍNDICE DE TABLAS

1.	Ejemplos de holarquias según Turnbull	30
2.	Relación entre jerarquías, estructuras edge y holarquías	38
3.	Diferencias entre los dos principales enfoques de fusión holónica	56
4.	Parámetros del sistema por tipo de aplicación	101
5.	Pruebas realizadas y resultados obtenidos para el caso A.	116
6.	Pruebas a realizar para el Caso B.	123
7.	Resultados de las pruebas para el Caso B	125
8.	Análisis de los resultados de las pruebas con rechazo parcial de H_0	128
9.	Pruebas a realizar para el Caso C.	134
10.	Análisis de los resultados de las pruebas con rechazo parcial de H_0	136
11.	Resultados de las pruebas del caso C.	137
12.	Calidad de la información de cada agente para cada campo	152
13.	Tabla MEJOR	153
14.	Tabla MEJORES	153
15.	Tabla MENSAJES-RESTANTES	154
16.	Evolución de la tabla “MEJOR”	155
17.	Evolución de la tabla “MEJORES”	156
18.	Evolución de “MENSAJES-RESTANTES”	157
19.	Significado del valor TT	199
20.	Modelo de matriz de dependencias de campos	202
9.3.	Composición de los campos caso B.	205
22.	Valores posibles de cada campo caso C	211

CAPÍTULO 1

INTRODUCCIÓN, MOTIVACIÓN Y OBJETIVOS

El aumento sostenido de la conectividad entre los sistemas y el advenimiento de la Internet de las Cosas (en inglés, *Internet of Things* o IoT), han conducido al uso de cada vez más información proveniente de fuentes de datos heterogéneas. Este creciente volumen de datos que se maneja hace que su calidad se vuelva algo crítico para facilitar la toma de decisiones (Janssen *et al.*, 2017);(O'Reilly, 1982). La calidad de los datos (*Data Quality*, DQ) puede ser vista como un conjunto de propiedades o dimensiones que describen las bondades de la información producida por una fuente de información (DeLone & McLean, 1992). Los términos calidad de los datos y calidad de la información (*Information Quality*, IQ) serán usadas en este trabajo de manera indistinta como términos equivalentes, tal como hacen otros autores (Knight & Burn, 2005). La IQ suele comprender las dimensiones de precisión y actualidad (*timeliness*), completitud (*completeness*), consistencia (*consistency*), relevancia (*relevance*) y adecuación al uso (*fitness for use*)(Miller, 1996; Wand & Wang, 1996), aunque otras investigaciones han considerado dimensiones adicionales, según los distintos criterios usados para evaluarla.

La medición de la IQ presenta varios desafíos. Por una parte, es básicamente subjetiva y variable en el tiempo: la importancia percibida de la calidad de un dato varía con el paso del tiempo para el usuario. Por otra parte, muchas de las dimensiones mencionadas en la bibliografía son extremadamente difíciles de medir (Knight & Burn, 2005), por lo que se persigue definir la IQ de forma que pueda ser cuantificada. Finalmente la IQ puede

tener varias representaciones (visuales, numéricas enteras discretas, numérica continuas, variables lingüísticas, etc.), siendo algunas de ellas de mayor comprensibilidad para las personas. Dado que existe cada vez mayor interacción entre personas y máquinas, resulta atractivo tener una representación que sea usable por personas y computadoras y que además no requiera de demasiados recursos para evaluarla.

En este trabajo de investigación se consideran esenciales dos dimensiones de la IQ por ser típicas de las comunicaciones entre personas: vaguedad e incertidumbre. La vaguedad y la incertidumbre son consideradas como defectos de la información por algunos autores (Frank, 2008; Smith, 2011) y están relacionadas con la indeterminación. En el contexto de este trabajo “indeterminación” se refiere al grado de conocimiento que se tiene sobre las consecuencias inmediatas de un mensaje.

Una forma de manejar dicha vaguedad e incertidumbre de la información, y mejorar los resultados de la toma de decisiones basadas en la misma, es usando fusión de información (*Information Fusion*, IF). En el caso especial de sistemas que incorporan personas, la IF es el método “natural” que éstas utilizan para manejar esa vaguedad e incertidumbre, ya que consultan distintas fuentes y luego procesan la información obtenida para alcanzar resultados con mayor certeza o de menor vaguedad (Nakamura *et al.*, 2014; Raol, 2015). Otra ventaja del enfoque de IF para manejar la indeterminación es que los participantes pueden razonar considerándola de distintas formas. Es por lo tanto de especial interés para sistemas abiertos que no garanticen la composición y características de sus miembros. Es decir, IF se realiza, entre otros motivos, para reducir el impacto negativo de las imperfecciones de la información. Al reducir el impacto negativo de esas imperfecciones, el receptor va a poder tomar mejores decisiones a la luz del contenido semántico asignado al mensaje recibido.

En lo que sigue, se entenderá por sistema un conjunto de componentes que tienen relaciones entre sí y forman un todo unificado, y que se distingue además de su entorno (Meer & Koppen, 2005a).

En el mundo real, los agentes (personas, máquinas, organizaciones, etc.) que intervienen en esas comunicaciones suelen tener recursos limitados (energía limitada, costo de la comunicación, riesgo táctico de realizar un número alto de comunicaciones, potencia de cómputo limitada, ancho de banda limitado, etc.), por lo cual las redes que se puedan

formar al sumarse agentes no siempre permiten trabajar sin restricciones en el número de mensajes que envían y/o reciben sus agentes miembros o en el tiempo empleado para dichas comunicaciones. Ejemplos de ello son: a) una red que utiliza equipos móviles (por ejemplo teléfonos móviles y tabletas) donde, o bien la batería se agota, o bien el saldo (crédito) para usarlo (en términos de llamadas o de tráfico de datos) se termina; b) redes sometidas a límites de tiempo de respuesta estrictos y en las que éste es un factor crítico (por ejemplo, en el manejo de incidencias producidas por desastres naturales o actos terroristas); o c) redes de uso militar con alta rotación de sus componentes o de tipo EDGE (Alston *et al.*, 2005b)(Alberts & Hayes, 2005) .

En forma genérica esas redes con limitaciones han sido llamadas redes de recursos limitados; las más comunes incluyen redes inalámbricas *ad hoc* (Milic & Malek, 2009), redes móviles *ad hoc* (MANETs) (Alston *et al.*, 2005b; Tavli & Heinzelman, 2006), redes de sensores inalámbricas (*Wireless Sensor Networks*, WSN) (Akyildiz *et al.*, 2002), redes tolerantes a demoras (*Delay Tolerant Networks*, DTN)(Alston *et al.*, 2005b) (Cao & Sun, 2013), redes vehiculares (VANets) (Alston *et al.*, 2005b)(Bujari *et al.*, 2019) (Karagiannis *et al.*, 2011) (Mujeeb Ur *et al.*, 2018) , *body area networks* (Chen *et al.*, 2011). Aplicaciones concretas de las redes anteriores son la toma de decisiones distribuidas, sistemas de recomendación en tiempo real y con geolocalización, etc. En dichas redes interesa alcanzar los objetivos usando los recursos limitados en forma eficiente.

Una topología común de estas redes es la denominada *peer-to-peer* (P2P) abierta, que permite que cualquier par de integrantes se comuniquen, sin imponer especializaciones en su estructura ni en sus elementos, y en las que éstos pueden además salir o entrar de la misma en cualquier momento.

Un sistema *peer-to-peer* es aquel en el que el cómputo se realiza compartiendo recursos (capacidad de proceso, ancho de banda, etc.) y servicios (Ou *et al.*, 2010) (Tutschku & Tran-Gia, 2005).

Un sistema P2P debe ser evaluado de acuerdo a la función que realiza ya que encontrar una métrica genérica es prácticamente imposible (Tutschku & Tran-Gia, 2005).

En estas redes (o sistemas) P2P abiertas(os) los integrantes suelen colaborar para lograr algún objetivo común. Se define un agente colaborativo como una colección de sensores (por *hardware* o *software*, reales o virtuales), actuadores, controladores, personas,

etc. que se comunican entre sí para alcanzar, en forma más o menos autónoma, una meta común. Se considerará a la red P2P como formada por agentes colaborativos.

Un espacio inteligente (*Intelligent Space*, IS) es considerado aquí como un sistema ambiental (*environment system*) que da apoyo a personas con información y en forma física, por ejemplo, usando monitores y robots (Gadeo-Martos *et al.*, 2011). En caso de que los componentes que forman un agente colaborativo tengan existencia física en un cierto espacio, y siempre que dicha existencia impacte en el comportamiento de otros componentes del mismo, se puede ver que se cumple la definición de espacio inteligente, gracias al agente colaborativo y a sus componentes. A un espacio (físico o virtual) poblado por agentes colaborativos lo llamamos espacio inteligente generalizado (*Generalized Intelligent Space*, GIS). Dos propiedades deseables de un IS (y de un GIS) son la flexibilidad y escalabilidad. En la naturaleza existen estructuras organizativas que son altamente flexibles, resistentes a los cambios que pueda haber en su entorno y que escalan muy bien al crecer el número de integrantes: es el caso de los holones y las estructuras holónicas (holarquías). Arthur Koesler introdujo el concepto de holón para referirse a entidades que pueden ser consideradas un todo y una parte a la vez: “que se comportan en parte como un todo o totalmente como partes, dependiendo de la forma en que se las considere”. Esas entidades pueden ser seres vivientes u organizaciones sociales (Koestler, 1982). Más formalmente, las holarquías son jerarquías dinámicas que se caracterizan por hacer uso eficiente de sus recursos, ser robustas frente a cambios en su entorno y ser más flexibles que una jerarquía debido a que explotan la autonomía de sus unidades funcionales (Horling & Lesser, 2004) y que están formadas por holones (Botti & Giret, 2008). Se puede decir que las holarquías garantizan la estabilidad, predictibilidad y optimización global del control jerárquico (Valencia-Jiménez & Fernández-Caballero, 2008). Ejemplos de holones naturales son: una célula, un hígado humano o el sistema solar; un sistema multi-agente autónomo, un equipo de trabajo o una fábrica organizada holónicamente podrían ser ejemplos de holones artificiales. Los holones pueden ser vistos como agentes abstractos con un cuerpo (es decir, agentes que tienen una parte física que interactúa con su entorno). Se suele decir que los holones tienen estructuras auto-similares, de tipo fractal (Paggi & Alonso, 2010). Un agente abstracto es un agente que puede ser considerado como formado por otros agentes. Giret y Botti definen; “Un agente abstracto

es un sistema de software con una entidad única, que está ubicado en un entorno, que, como un todo, percibe el entorno (tiene entradas sensibles al entorno). A partir de esas percepciones determina y ejecuta acciones en una forma autónoma y flexible, reactiva y proactiva. Desde un punto de vista de la estructura, un agente abstracto puede ser un agente (entidad atómica) o puede ser un sistema multi-agente (con una entidad única) formado de agentes que no son necesariamente homogéneos. Un Sistema multi-agente está formado por dos o más agentes abstractos que interactúan para resolver problemas que están más allá de las capacidades individuales y el conocimiento individual de cada agente abstracto.” (Botti & Giret, 2008).

Así como los cristales de hielo, que son un análogo físico de los holones por tener asimismo una estructura fractal, se forman al reducir la temperatura de cierta forma, y dado que tradicionalmente se sostiene que la incertidumbre tiene un efecto desorganizador en las organizaciones humanas (González, 2009), resulta interesante ir más allá e investigar el problema dual de si el aumento de la calidad de la información tiene un efecto similar en los GIS, produciendo holarquías. Es decir, si ese aumento de calidad de la información lleva a la formación de estructuras auto-similares, como existen en los sistemas caóticos (Thiéart & Forgues, 1995).

Entre las aplicaciones de los GIS se encuentran los sistemas de *soft sensors* distribuidos, de gran difusión en la industria, motivo por el cual en este trabajo se incluye un caso de estudio relacionado con ellos.

1.1. Motivación

La expansión de la conectividad entre distintos tipos de objetos ha permitido que cada vez más variedades de agentes (mecánicos o humanos) colaboren formando redes que buscan alcanzar un fin; esos sistemas así formados son, a semejanza de la web, abiertos, dinámicos, de procesamiento distribuido y generadores de enormes cantidades de datos. Ese es el caso de IoT. Dada la gran variedad de agentes que esto conlleva, tener un marco para medir la IQ es fundamental para aplicaciones distribuidas donde todos componentes pueden dar respuestas de calidad variable.

Muchos de los equipos que integran esas redes pueden tener baterías de duraciones muy

limitadas (para que sean pequeños o livianos), escaso ancho de banda para comunicarse o pueden ser lentos por poseer escasa potencia de cómputo. Estas limitaciones de los recursos pueden afectar mucho la toma de decisiones si no se los administra racionalmente.

Este trabajo se relaciona con el uso potencial de las redes de recursos limitados en el apoyo a la toma de decisiones en las que se utiliza el concepto de IQ de forma de proveer buenas opciones, con costos aceptables (asociando los costos a mensajes enviados, energía usada, etc.).

Dado lo abierto y dinámico de las redes generadas en la IoT se asume que los agentes no tienen una organización específica pero sí que pueden comunicarse libremente entre sí, por lo que se los considera formando una red P2P plana, cuyos individuos poseen recursos limitados. Ejemplos típicos de esto son a) el caso de una comunidad cuyos integrantes colaboran para hallar la respuesta más adecuada a una consulta que se les plantea desde afuera (tal como un sistema de recomendación de hoteles u atracciones turísticas) donde los usuarios (*peers* de la red) pueden dar en tiempo real su opinión (respuesta) a una consulta de alguien sobre la conveniencia de una opción usando su *smartphone* que tiene batería y tal vez número de mensajes limitados; b) un escenario donde un grupo de personas está ayudando a manejar una catástrofe y se comunican usando dispositivos que tienen limitaciones (batería, duración de las transmisiones, etc.) y donde se pueden incorporar nuevas personas en cualquier momento (Groen *et al.*, 2017); o c) un escenario de un grupo realizando maniobras militares en terreno hostil, donde la cantidad de mensajes está limitada por consideraciones de seguridad, por ejemplo, y donde ningún integrante puede considerarse imprescindible para la misión.

En este trabajo, al referirse a agentes colaborativos (o más brevemente, “agentes” o “*peers*”), no se está suponiendo que todos son del mismo tipo ni que son todos personas o máquinas. En el caso de una red de sensores por *software* (*soft sensors*), muy populares en el modelo de industria 4.0, cada uno de ellos va a ser calculado por un agente que puede ser humano o no. Los agentes a su vez colaboran y así se pueden generar *soft sensors* de segundo, tercer nivel o sucesivos.

Dado que la información y su calidad son usadas para la toma de decisiones, es interesante la simplificación que aporte el que se pueda representar su calidad por un solo número evitando así una toma de decisiones basada en múltiples criterios. Por otra

parte, el manejo de la vaguedad e incertidumbre es importante en el diseño de espacios inteligentes, principalmente porque la vaguedad es típica de las comunicaciones entre humanos y para que la comunicación con un espacio de ese tipo resulte útil y cómoda a las personas deberá incorporarse el manejo de ese tipo de expresiones. La métrica de calidad a usar deberá incluir una estimación de dicha vaguedad e incertidumbre.

A pesar de la extensa teoría que existe para razonar en base a vaguedad e incertidumbre (teoría de las probabilidades, teoría de las posibilidades, lógica difusa, etc. – ver por ejemplo (Fagin & Halpern, 1994; Halpern, 2003) el uso de la IF parece lo más adecuado ya que cada agente puede razonar y manejar la vaguedad e incertidumbre en cualquiera de dichas formas y dar la respuesta (incluyendo su calidad) que corresponda en lugar de pretender calcularlas algorítmicamente como cuando se consideran todos los agentes homogéneos en la forma de manejar dichas vaguedades e incertidumbre y en cómo calcular las respuestas a partir de las entradas.

Como se dijo, las redes que se forman tienden a ser de elementos heterogéneos, lo que no asegura que todos manejen la indeterminación ni que realicen la IF de manera similar. Por esto es atractivo obtener un modelo en el que la IF quede solo especificada a muy alto nivel.

Por otra parte, las mencionadas redes podrán ser parte de otras redes o estar en entornos cambiantes. Interesan, por lo tanto, los modelos de sistema que permitan la adaptación, por ejemplo, porque se auto-organizan en forma de holarquías. Se ha encontrado que cuando la racionalidad es limitada (capacidad de procesamiento de información y de comunicaciones, típico de los equipos pequeños usados en la IoT), es acotada y se enfatiza la velocidad de adaptación al entorno, los sistemas formados por bloques ‘grandes’ tienen mejor desempeño que los formados por bloques ‘pequeños’ (Valckenaers & Van Brussel, 2005), de ahí la importancia de los holones (entendiendo por bloque a un conjunto de agentes que actúan coordinadamente).

Tómese como ejemplo una red de sensores que determina si el nivel de ozono a nivel del suelo es peligrosamente alto, tal como se describe en la sección 6.5. Un nivel de concentración de ozono por encima de cierto umbral conocido es peligroso para la salud humana y afecta actividades como trabajos agrícolas, turismo, etc. Por tanto es necesario tener un sistema preciso de predicción indirecta de esa concentración (midiendo otras

variables atmosféricas) para enviar las alertas necesarias antes de que alcance niveles peligrosos.

Para hacer que esas predicciones sean más significativas se puede realizar la medición en una gran superficie geográfica y usar los datos proporcionados por una cantidad importante de sensores distribuidos en toda ella. Estos sensores poseen escasa capacidad de cómputo y baterías de no gran duración, siendo de bajo costo a los efectos que su pérdida o rotura no sea un evento importante.

Algunos de los sensores pueden dañarse por estar cerca del suelo, por lo que la calidad de su información no está asegurada. Se comprende que realizar el mantenimiento de sus baterías puede requerir tiempo y trabajo significativo cuando la dispersión geográfica es alta, por lo que se hace necesario tener comunicaciones muy eficientes de forma de prolongar la carga o vida útil de esas baterías y que es imprescindible considerar la calidad de la información provista por cada sensor para contemplar los casos de mal funcionamiento, así como una forma de operar que maneje “suavemente” las situaciones en que sensores dejan de funcionar o desaparecen. En entornos del tipo del que se acaba de describir (*soft sensors*), en los que es necesario realizar fusión de información entre agentes para lograr una meta común siendo los recursos limitados, es donde tiene especial aplicabilidad el modelo descrito en esta Tesis.

1.2. Objetivos

Sea un conjunto de elementos que se conectan entre sí sin ninguna estructura predefinida y sin requerir ninguna especialización de parte de ellos, de forma de alcanzar colectivamente algunas metas, como por ejemplo responder a una consulta que plantea un agente externo Ω . En ese caso los elementos formarían un espacio inteligente generalizado (ya que los mismos pueden o no existir en el mundo físico) que ayudaría a Ω . El conjunto es abierto y en cualquier momento pueden ingresar o salir integrantes no presuponiéndose nada sobre su homogeneidad.

Estos elementos, en el mundo real, tienen recursos (tiempo, cantidad de mensajes disponibles, energía, etc.) limitados, por lo que se hace necesario manejar las comunicaciones eficientemente, y pueden generar información con ciertas “imperfecciones” que afectan su

calidad, tales como la incertidumbre (por ejemplo debido a un error de medición) o vaguedad (típica de cuando intervienen personas, como en el caso de expresiones *fuzzy* del tipo “hombre alto” o “distancia larga”). Otras facetas de la calidad a tener en cuenta son su confiabilidad, credibilidad, eficiencia y confidencialidad, que son importantes en ambientes donde todos se pueden comunicar con todos.

Al existir una gran cantidad de integrantes, los volúmenes de datos generados en la red son inmensos, tal como ocurre en la IoT y Big Data. Estos altos volúmenes de información de distintas calidades dificultan la toma de decisiones. Dichas calidades son una metadata que puede tener distintas representaciones, algunas muy sencillas de interpretar por una persona (por ejemplo: un diagrama de telaraña) pero no por una máquina, y recíproco, por lo que interesa una representación que sea utilizable por personas y máquinas y que simplifique la toma de decisiones.

Dado que se supone que algunos de los elementos pueden ser personas es natural pensar que esas imperfecciones se tratan de sobrellevar haciendo fusión de la información original “defectuosa” con la de otras fuentes. La IF permite no considerar los formalismos de razonamiento con incertidumbre y vaguedad (probabilidades, teoría de la posibilidad, lógica *fuzzy*, etc.) que podrían ser inusables en el caso de agentes heterogéneos.

Se entenderá por **éxito** al evento de que la respuesta obtenida por Ω usando el sistema tenga una calidad mayor o igual que la que tenía originalmente Ω . El éxito será llamado **estricto** cuando la calidad sea mayor que la de Ω . Como se entiende, los éxitos interesantes son los estrictos, ya que denotan cuando el uso del sistema resultó útil para Ω .

Considerando todo lo anteriormente expuesto se comprende que interesa encontrar un modelo inteligente de comunicaciones en redes P2P, en las que sus elementos tienen recursos limitados y donde se realiza IF para mejorar la calidad de la información que se intercambia la cual puede contener vaguedad e imprecisión, para alcanzar una meta (tomar una decisión). Se buscará validar o refutar dos hipótesis sobre dicho modelo, a saber:

- **H1** - el modelo permite mejorar el rendimiento (medido considerando los éxitos estrictos obtenidos y mensajes consumidos).
- **H2** - el modelo posibilita la auto-organización de la red de forma que permita reaccionar con flexibilidad a los cambios en su entorno y en su composición, manteniendo su estabilidad y predictibilidad, más específicamente, que conduzca a la formación de estructuras holónicas.

1.3. Aportaciones

Las principales aportaciones del presente trabajo son:

- **A1.** Un modelo inteligente de comunicaciones en redes P2P que, usando IF, permite mejorar el rendimiento de la red (medido como éxitos obtenidos y mensajes consumidos) en escenarios con recursos limitados. Dicho modelo se basa en la definición de una métrica de calidad de la información, de una arquitectura parametrizable que permite articular el funcionamiento del modelo en tres grandes etapas o fases y de una serie de comportamientos que adoptan los agentes de la red en cada momento.
- **A2.** La demostración teórica de la formación de estructuras holónicas en la organización de los agentes a partir del propio funcionamiento del modelo propuesto.

1.4. Contribuciones

Esta Tesis resume un conjunto de investigaciones relacionadas con el manejo de información con vaguedad e incertidumbre, y en forma más general con problemas en su

calidad, la colaboración entre agentes para formar sistemas complejos y las estructuras holónicas. Los resultados han sido divulgados en una serie de conferencias y artículos científicos que se detallan en el APÉNDICE: Contribuciones.

Otra contribución es el código fuente de los programas desarrollados, que ha sido hecho de acceso libre, en la dirección <https://github.com/hpaggi/programas-tesis>.

1.5. Organización del documento

Este trabajo está organizado como sigue. A continuación, en el Cap. 2 se describe el estado de la cuestión respecto los principales tópicos tratados aquí. En el Cap. 3 se plantea el problema y se establecen las preguntas de investigación (o *research questions*) que guiaron esta investigación. Luego, en el Capítulo 4 se hace una revisión de las soluciones dadas en la bibliografía a estas preguntas y se muestra la necesidad y utilidad del enfoque aquí planteado. Dicho enfoque junto con otros datos de la solución propuesta son dados en el Capítulo 5. El trabajo experimental realizado a través de los distintos casos de estudio, que muestran estadísticamente las ventajas de este modelo sobre el simple, así como la descripción paso a paso de la formación de los holones en el sistema en un ejemplo, son mostrados en el Capítulo 6. En el capítulo 7 se muestra que un sistema que implemente este modelo puede generar holones y cómo, con su funcionamiento, la probabilidad de que las estructuras generadas no sean holones tiende a 0 al crecer el número de agentes del sistema. El Capítulo 8 incluye las conclusiones globales y el Capítulo 9 comenta algunas líneas de trabajo futuro que podrían extender este trabajo. La bibliografía completa el documento.

CAPÍTULO 2

ESTADO DE LA CUESTIÓN

A continuación se describe el estado de la cuestión en los principales aspectos que inciden en el problema planteado, a saber: calidad de la información, indeterminación, vaguedad e incertidumbre, redes de recursos limitados, sistemas *peer-to-peer*, objetos colaborativos y espacios inteligentes, *sensemaking*, *soft sensors*, fundamentos de estructuras holónicas, fusión de información y auto-organización.

2.1. Calidad de la información

Por “calidad”, el diccionario de la Real Academia define “propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor”. En este trabajo calidad de los datos y calidad de la información (en inglés, *Information Quality*, IQ) serán usados como términos equivalentes, tal como otros autores hacen (Knight & Burn, 2005). La calidad de la información (IQ) puede ser vista como un conjunto de propiedades (o dimensiones) que describen las características deseables (exactitud, significación, etc.) de la misma producida por un sistema de información (DeLone & McLean, 1992). Esta calidad suele comprender las dimensiones de precisión (*precision*), actualidad (*timeliness*), completitud (*completeness*), consistencia (*consistency*), relevancia (*relevance*) y adecuación al uso (*fitness for use*) (Miller, 1996; Wand & Wang, 1996). Los investigadores han considerado también otras dimensiones adicionales, si bien aún no han llegado a un consenso,

por lo que se han usado distintos marcos (*frameworks*) para evaluarla. Una descripción esquemática de doce de ellos se encuentra en Knight & Burn (2005).

Uno de los aspectos más importantes de la IQ es que influirá en la calidad de las decisiones tomadas a partir de la información de que se dispone Janssen *et al.* (2017); O'Reilly (1982).

La medición de la IQ presenta dos desafíos. Uno, es que es, básicamente, subjetiva y varía con el tiempo (la importancia que tiene la calidad de un dato para su usuario puede variar con el tiempo); el otro, es que muchas de las dimensiones mencionadas en la bibliografía son extremadamente difíciles de medir Knight & Burn (2005), por lo que es interesante definir la IQ de manera que pueda ser medida. Al respecto hay pocos antecedentes; en el caso de IQ para recuperación de información (*information retrieval*), se tienen los trabajos de Zhu y Gauch (Zhu & Gauch, 2000), Naumann y Rolker (Naumann & Rolker, 2000), Eppler y Manenzmayer (Eppler & Muenzenmayer, 2002), Knight y Burn (Knight & Burn, 2005).

La calidad de la información puede tener varias representaciones (visuales, numéricas enteras discretas, numérica continuas, variables lingüísticas, etc.), siendo algunas de ellas de mayor usabilidad para los humanos. Un estudio detallado de ello se puede ver en (Price, 2011). Poder medir la calidad de la información posibilitará la selección de la fuente de información con la mejor calidad. En el caso del modelo propuesto en esta Tesis no se consideró únicamente la usabilidad por parte de humanos, sino que se buscó que fuese la representación más simple, manejable por humanos y máquinas a la vez, y que, además, fuese sencilla de calcular y, así, no impusiese una sobrecarga importante de cómputo al agente.

2.2. Indeterminación, vaguedad e incertidumbre

Dos facetas de la calidad de la información interesan especialmente en este trabajo: la vaguedad y la incertidumbre; es decir, cuán exacto o confiable es un ítem de datos y cuán vago puede ser. La vaguedad y la incertidumbre son dos componentes muy importantes en la calidad de la información. Ambas están relacionadas con la indeterminación. En el contexto de este trabajo, la indeterminación se refiere al grado de conocimiento que hay

sobre las consecuencias inmediatas de un mensaje. Como dice Novák, la incertidumbre y la vaguedad forman dos facetas complementarias de un fenómeno más general que se puede llamar indeterminación (Novák, 2005). La indeterminación (o sea, incertidumbre y vaguedad) implica un grado de creencia en la proposición comunicada y, a la vez, sólo una tendencia a actuar: citando a Smith, “cuando un término es familiar, puede ser usado sin que se pregunte “¿qué significa eso?”... un grado de creencia en que si la proposición P [es verdadera] implica una tendencia a actuar como si P [es verdadera]”. Dicha indeterminación se refiere a cuánto se sabe sobre las consecuencias de recibir un mensaje. Por lo tanto interesa la vaguedad e incertidumbre en la medida que puedan afectar a las respuestas y a las creencias de los agentes, como en (Cochez, 2012). Para un análisis detallado de los distintos aspectos de la indeterminación se refiere al lector a (Bossé & Solaiman, 2016). Se pueden comparar dos medidas de calidad de información sintetizando sus dimensiones (vaguedad, incertidumbre y otras relacionadas con la confiabilidad, credibilidad y eficiencia) en un único valor, que es a lo que se llama con el nombre de calidad.

La incertidumbre está relacionada principalmente con el error o imprecisión asociado a los datos y la vaguedad es una característica típica del lenguaje natural (por ejemplo, en la frase “trabajo extenso”, ¿cuántas páginas significa “extenso”?).

De las diferentes formas posibles de modelar la vaguedad (usando lo que Sutton llama “teorías de grado” (Sutton, 2013) se optó en esta Tesis por la teoría de conjuntos difusos de Zadeh (*fuzzy set theory*) (Zadeh, 1973), lo cual es usual (ver por ejemplo (Ekel *et al.*, 2016; Flores *et al.*, 2009) (Solano-Aragón & Garza, 2009). Por una parte, uno de los componentes de la información vaga es su vaguedad, su *fuzziness* (su dispersión), es decir que la calidad es más pobre cuanto más vaga es la información. Esta idea es impulsada por ejemplo por Bardossy *et al.* (Bardossy *et al.*, 1993; Hsu & Chen, 1996), diciendo que el mejor de dos números difusos es el menos vago. Por ejemplo, la dispersión de una respuesta expresada como un número *fuzzy* R puede ser medida como el área debajo de

su función de pertenencia (Bardossy *et al.*, 1993), formalmente es (Zi-Xiao, 1984):

$$\begin{cases} H(R) = \int_{-\infty}^{+\infty} \mu_R(x) dx & \text{caso continuo} \\ H(R) = \sum_{x \in X} \mu_R(x) & \text{caso discreto} \end{cases} \quad (Ec. 1)$$

Otras formas de medirla se han creado. Por ejemplo, Yager (citado por (Ayyub & Klir, 2006)) propone (cuando R es finito):

$$H(R) = |X| - \sum_{x \in X} \mu_R(x) - \mu_{R^-}(x) \quad (Ec. 2)$$

donde R^- es el complemento de R y X el conjunto universal donde varía x . Esto mide la falta de distinción entre un conjunto *fuzzy* y su complemento, es decir, cuán *fuzzy* o borroso es.

Se han desarrollado además otras medidas de *fuzziness*, por ejemplo en (Delgado *et al.*, 1998; Ebanks, 1983)(Zi-Xiao, 1984) que se omiten por brevedad y por no ser relevantes para este trabajo. Esta idea de medir la calidad de un número según la dispersión de los valores que puede tener es la que se ha implementado en la métrica de calidad de la información propuesta en este trabajo (ver Cap. 5). Por otra parte, la incertidumbre, que se refiere a la no especificidad propia del campo de datos en consideración, es decir, al número de alternativas posibles para dicho campo (conjunto de distintos valores que puede tomar) según el receptor del mensaje, es modelada tradicionalmente con la teoría de la probabilidad. La calidad del dato R entonces incorporará su vaguedad y su incertidumbre:

$$Q(R) = Q(f(H(R), g(u(R))) \quad (Ec.3)$$

donde f y g son ciertas funciones decrecientes de la dispersión H y la incertidumbre u , por ejemplo:

$$Q(R) = \frac{1}{(1+H)(1+u)} \quad (Ec.4)$$

El manejo de la vaguedad y de la incertidumbre es importante en el diseño de espacios inteligentes, principalmente porque la vaguedad puede darse en las comunicaciones entre agentes (personas, sistemas, etc.), y para que la comunicación de dichos agentes con ese espacio resulte útil y cómoda a ellos, deberá incorporarse el manejo de ese tipo de

expresiones. La incertidumbre por su parte señala la certeza que tiene el agente sobre la información lo cual es básico para saber cómo considerar, o qué importancia darle, a dicha información.

2.3. Redes de recursos limitados

Si bien las posibilidades de comunicación entre personas y cosas ha ido en aumento, de forma que casi todo se comuniquen con casi todo, llevando a tener redes cada vez más complejas, debido a una serie de factores (energía limitada, costo de la comunicación, riesgo táctico de realizar un número alto de comunicaciones, potencia de cómputo limitada, ancho de banda limitado, etc.) no siempre esas redes pueden trabajar sin limitaciones en el número de mensajes ni tiempo. Ejemplos de ello son: a) una red usando equipos móviles (por ejemplo celulares y tabletas) donde o bien la batería se agota o bien el saldo (crédito) para usarlo se termina (lo que se aplica tanto a comunicación de personas o de cosas); b) redes sometidas a límites de tiempo de respuesta estrictos: cada comunicación consume tiempo y, en algunos casos, ese es un factor crítico (por ejemplo, en el manejo de incidencias producidas por desastres naturales o actos terroristas); c) redes de uso militar con alta rotación de sus componentes¹ o de tipo *edge*² (Alberts & Hayes, 2005).

De forma genérica, esas redes con limitaciones han sido llamadas redes de recursos limitados; las más comunes incluyen redes inalámbricas *ad hoc* (Milic & Malek, 2009), redes móviles *ad hoc* (MANETs) (Heinzelman, 2006; Alston *et al.*, 2005b; Bellavista *et al.*, 2018; Mujeeb Ur *et al.*, 2018) redes de sensores inalámbricas (WSN) (Akyildiz *et al.*, 2002; Habib *et al.*, 2018)(Zhong *et al.*, 2018), redes tolerantes a demoras (DTN *Delay Tolerant Networks* , por sus siglas en inglés) (Alston *et al.*, 2005a; Cao & Sun, 2013)(Jiagao *et al.*, 2018), redes vehiculares (VANets, *Vehicular Ad hoc Networks*) (Bujari *et al.*, 2019; Karagiannis *et al.*, 2011; Mujeeb Ur *et al.*, 2018), *body area networks* (Chen *et al.*, 2011).

¹Ya que por ser militar suponemos que la cantidad de recursos son medidos (batería, si el equipo es móvil y cantidad de mensajes debido a que se presume que a más mensajes más posibilidades de ser espiadas). La rotación de integrantes aumenta el efecto de agentes que desaparecen del sistema en cierto momento.

²El sentido de “*edge*” en este contexto se relaciona con las estructuras descentralizadas de comando y control donde un elemento puede acceder a una red de información global (*Global Information Grid*), tal como Alberts lo define en (Alberts & Hayes, 2005). Las estructuras *edge* permiten tantas comunicaciones entre sus componentes que pueden llevar a que los recursos de comunicación resulten escasos si no se usan inteligentemente

Aplicaciones concretas de las redes anteriores son la toma de decisiones distribuidas, sistemas de recomendación en tiempo real y con geolocalización, etc. En dichas redes interesa alcanzar los objetivos usando los recursos limitados en forma óptima.

Según Nakamura (Nakamura *et al.*, 2014), se puede usar información local (es decir la que tiene el agente y la información de su entorno) para obtener algoritmos muy eficientes en una red de recursos limitados. En base a ello, como se verá en secciones sucesivas de esta Tesis, se comparan los resultados obtenidos usando el modelo inteligente propuesto en la Tesis, que busca ahorrar recursos en redes donde éstos son limitados, con los resultados obtenidos en redes que no consideran dicho comportamiento inteligente.

2.4. Sistemas *Peer-to-Peer*

2.4.1. Generalidades

Un sistema, en el contexto de esta Tesis, se puede definir como un conjunto de componentes que tienen relaciones entre sí y forman un todo unificado, y que se distingue a sí mismo de su entorno (Meer & Koppen, 2005a). A su vez, un sistema *peer-to-peer* (P2P) es un sistema en el que el cómputo se realiza a través de la compartición de recursos (capacidad de proceso, ancho de banda, etc.) y servicios (Ou *et al.*, 2010)(Tutschku & Tran-Gia, 2005). Un sistema P2P debe ser evaluado de acuerdo a la función que realiza ya que una métrica de evaluación genérica es casi imposible (Tutschku & Tran-Gia, 2005). Un fenómeno típico e indeseado de algunos de estos sistemas es el *churn* (llegada y partida continua de los participantes de la red P2P), por lo que se suele tratar de evitarlo.

Las redes P2P pueden ser planas o jerárquicas. Una red plana es aquella en que las funcionalidades y carga están distribuidas uniformemente entre los nodos integrantes (Vu *et al.*, 2010a) mientras que en una jerárquica los servidores (nodos) de un nivel actúan como clientes de los servidores (nodos) de un nivel superior (por ejemplo en la red DNS).

Hay dos tipos de salidas del *peer* de la red. Se dice que se produce un *graceful leaving* cuando un *peer* avisa que se va y transfiere sus recursos a otros *peers*, que no es el caso del modelo desarrollado en esta Tesis. En el *ungraceful leaving* el *peer* que se va no notifica su partida ni transfiere recursos a sus vecinos (por ejemplo por falla o falta de energía),

y es el que corresponde al modelo presentado en esta Tesis ya que se pretende modelar detenciones no planificadas en las comunicaciones, como sucede cuando hay una baja en un grupo militar en acción o cuando se descompone un sensor en un sistema de sensores distribuidos.

Según Vu *et al.* (Vu *et al.*, 2010a) hay toda una gama de arquitectura de los sistemas P2P, que van desde las que admiten servidores que centralizan el servicio a otras totalmente descentralizadas donde eso no ocurre, pasando por modelos híbridos donde los nodos son organizados en dos capas: la capa superior de servidores y la inferior de nodos comunes. A continuación se describen algunas de estas arquitecturas:

a) Sistemas P2P centralizados

En este caso, para solicitar asignación de recursos, un *peer* manda mensajes al servidor central para determinar las direcciones de los *peers* que contienen los recursos deseados (por ejemplo Napster). Una vez que el *peer* tiene los recursos deseados se puede comunicar directamente con otros *peers* sin ir más a través de los servidores.

b) Sistemas P2P descentralizados

En los sistemas P2P descentralizados, los *peers* tienen iguales derechos y responsabilidades. Los sistemas descentralizados pueden ser a su vez clasificados en planos o jerárquicos. En los planos la funcionalidad y carga está distribuida uniformemente entre todos los nodos. La mayoría de los sistemas descentralizados existentes son no jerárquicos. Por otro lado los jerárquicos tienen ventajas naturales como el aislamiento de las fallas y de la seguridad, uso efectivo de las memorias *caché* y uso efectivo del ancho de banda, etc. (Vu *et al.*, 2010a). Según su topología, los sistemas descentralizados pueden ser estructurados o no estructurados (desestructurados). En los desestructurados cada *peer* es responsable de sus propios datos y lleva registro de los vecinos a quien puede reenviar consultas. No hay garantía de que las respuestas sean completas (a menos que se busque en toda la red) ni del tiempo de respuesta. Ejemplos de este modelo son las redes Gnutella (Frankel & Pepper, 2000) y FreeNet (Clarke *et al.*, 2001). En los estructurados, en cam-

bio, cada usuario es responsable de una parte específica del contenido en la red. Estas redes utilizan funciones de *hash* distribuido y asignan valores a cada contenido y a cada usuario en la red. Después siguen un protocolo global en la determinación de qué usuario es responsable de qué contenido. Una cuestión clave en los sistemas desestructurados es la determinación de los vecinos (o sea, a quién consultar). La lista de los vecinos a quién consultar va cambiando según cambia el tipo de consultas con el tiempo. A esos sistemas se les llama sistemas reconfigurables (por ejemplo la red BestPeer (Wee Siong *et al.*, 2002)). Por otra parte, en los estructurados, los datos (recursos) son colocados bajo el control de estrategias predefinidas (por ej. una tabla de datos *hash* distribuida). Es decir, hay una correspondencia entre los datos y los *peers*. Un ejemplo son las redes Chord (Antonopoulos, 2010) y Pastry (Rowstron & Druschel, 2001). En los sistemas P2P híbridos no hay servidores; sin embargo, algunos nodos más potentes que otros, pueden ser elegidos para actuar como servidores de otros. Esos nodos son llamados super nodos (*super peers*). Obsérvese que, si bien la mayoría de la investigación en estas redes es sobre la aplicación de compartir datos, se pueden usar también para otras aplicaciones (por ejemplo compartir almacenamiento, ciclos de procesamiento, etc.).

b.1) Sistemas totalmente descentralizados

En las arquitecturas totalmente descentralizadas cada *peer* tiene la misma responsabilidad y derechos. Es más, no hay servidores centralizados ni otros mecanismos auxiliares (supernodos) para coordinar las operaciones entre los *peers*. Estos sistemas se pueden clasificar en no estructurados y estructurados según la estructura de la red subyacente. En los sistemas P2P no estructurados, el contenido de cada *peer* no tiene relación con la “estructura” de la red subyacente. Cada *peer* elige el contenido a guardar a voluntad. La ventaja es que puede acomodar una población de nodos creciente, ya que las consultas se pueden distribuir a un gran número de nodos dentro de un corto tiempo (Vu *et al.*, 2010b). La principal desventaja es que se inunda con la consulta a muchos nodos que no la pueden responder, lo que causa un gran tráfico de mensajes. Si bien esta estrategia es eficiente en cuanto al uso de ancho de banda, la escalabilidad de estos sistemas es problemática (Vu *et al.*, 2010b). Un ejemplo de esta arquitectura son las redes Gnutella, FreeNet y Freehaven. En los estructurados, hay un cierto mecanismo para determinar la

ubicación en la red, o sea que los archivos, si se piensa en la función de la red como de almacenamiento de datos, son colocados en ciertas ubicaciones (por ejemplo, usando la función de *hash* distribuida).

b.2) Propiedades de los sistemas descentralizados

En los sistemas descentralizados, la red tiene una fuerte inmunidad a la censura, a las fallas técnicas y a los ataques maliciosos. En los sistemas no estructurados, el encaminamiento de mensajes es alcanzado a menudo a través de *broadcasting* de mensajes lo que involucra a un gran número de *peers* en el proceso de búsqueda durante un corto tiempo. Para evitar que se produzca así una inundación de mensajes en la red, se le puede poner a los mensajes un tiempo de vida (*time to leave*) que se decrementa con cada reenvío y el mensaje se descarta al llegar a 0.

c) Sistemas híbridos P2P

En algunos sistemas híbridos hay peers que tienen mucha más capacidad de procesamiento que los otros (*super peers* o super nodos). Estos super nodos forman un nivel superior que tienen servicios parecidos a los servidores de un sistema centralizado; los *peers* comunes tienen muchos más servicios de parte de esos super nodos. Se debe tener presente que:

- Un supernodo no es tan potente como un servidor en un sistema centralizado.
- Un servidor en un sistema centralizado sólo ayuda a los *peers* a localizar los recursos sin compartir los recursos él mismo. En cambio, el supernodo coordina las operaciones de otros y además contribuye con sus recursos al funcionamiento de la red.
- Con la característica de auto reconfiguración cada *peer* en un sistema híbrido, se las arregla para conectarse con aquellos que potencialmente pueden beneficiar sus consultas posteriores. En suma, los sistemas híbridos tienen ventajas tales como optimizar la topología de la red, mejorar el tiempo de respuesta y ahorrar consumo de recursos y evitar un solo punto de falla.

2.4.2. Redes P2P en el contexto de la Tesis

Cuando se consideran redes de recursos limitados, se hace necesaria una forma de trabajo de agentes colaborativos que permita una reducción de los recursos usados comparándolo con una en el que simplemente se interconectan todos los componentes del agente.

En el caso de redes P2P, tener un marco para medir la IQ es fundamental para aplicaciones distribuidas donde todos componentes (máquinas o humanos) se pueden conectar libremente entre sí, generándose enormes cantidades de datos, formando sistemas que, a semejanza de la web, son abiertos, dinámicos, de procesamiento distribuido y con partes poco confiables. Ese es el caso de IoT.

Cuando los recursos son limitados, la toma de decisiones puede verse muy afectada por la falta de aquellos si no se los administra racionalmente. Esta Tesis se relaciona con el uso potencial de las redes de recursos limitados en la toma de decisiones, con el objetivo de proveer buenas opciones, con costos aceptables (asociando los costos a mensajes enviados, energía usada, etc.).

2.5. Objetos colaborativos y espacios inteligentes

Un objeto colaborativo es “una colección de sensores, actuadores, controladores u otros objetos colaborativos que se comunican entre sí para alcanzar, más o menos autónomamente, una meta común” (Banatre *et al.*, 2008). Los objetos colaborativos están asociados a aplicaciones de muy distintos tipos, desde robots trabajando en equipo a redes de sensores wireless. Aquí se ampliará la noción y se definirá un agente colaborativo como una colección de sensores (*hard* o *soft*, reales o virtuales), actuadores, controladores, personas u otros agentes colaborativos que se comunican entre sí para alcanzar, más o menos autónomamente, una meta común. Como se ve, una persona puede ser considerada un agente colaborativo. La denominación “agente” en lugar de “objeto” refuerza la idea de “más o menos autónomamente” a través del concepto de agencia característico de los agentes.

En el caso que los componentes que forman un agente colaborativo tengan existencia física en un cierto espacio y si dicha existencia impacta en el comportamiento de otros

componentes del mismo se observa que el agente colaborativo hace que ese espacio sea inteligente según la definición de Lee (Lee *et al.*, 2004): un “espacio inteligente” (IS, *Intelligent Space*) es un sistema ambiental (*environment system*) que da apoyo a las personas con información y en forma física, por ejemplo, usando monitores y robots (Gadeo-Martos *et al.*, 2011).

Asimismo un agente colaborativo tendrá, por definición, propiedades similares a las que un espacio físico debe cumplir para ser inteligente, propuestas por Mohan (citado por (Gadeo-Martos *et al.*, 2011a)):

- a) Ser diseñado para las personas y para facilitar las actividades humanas normales que tienen lugar en, o interactuando con, ellos.
- b) La información sobre el estado de componentes del agente y de los eventos que suceden en él son capturadas automáticamente por los propios componentes del agente.

A un espacio (físico o virtual) poblado por agentes colaborativos lo llamamos espacio inteligente generalizado. Por ejemplo, en el contexto de esta Tesis, cuando un conjunto de agentes de software interactúan entre sí para dar la mejor respuesta posible a un usuario que realiza una consulta, esos agentes conforman un espacio (virtual) inteligente generalizado.

Se generalizan aquí las funciones esenciales de un IS definidas en (Sasaki *et al.*, 2007) a los agentes y sensores definidos aquí así como los IS generalizados, por lo cual ellas serían:

- a) Observar el espacio (físico o virtual) en el que se desenvuelve el agente, usando sensores distribuidos del tipo adecuado (físicos, *por software* o virtuales).
- b) Extraer información útil de esos datos obtenidos y proveer servicios varios.

Otras propiedades de los IS son (Sasaki *et al.*, 2007) :

- a) El sistema debe tener una función que fusiona la información adquirida por cada sensor y que comparte con otros sensores eficientemente.
- b) El sistema debe tener flexibilidad y escalabilidad.

2.6. *Sensemaking*

La idea de *sensemaking* (o *sense-making*) fue introducida en el campo de la Information Science por Brenda Dervin (Dervin, 1998, 1983, 2003), en la interacción hombre-máquina por Russell, Stefik, Pirolli and Card (Russell *et al.*, 1993) y en los estudios organizacionales de Karl E. Weick (Weick, 1995, 1993, 1979). El concepto unifica ideas de la filosofía, sociología y ciencia cognitiva (en especial, psicología social) por lo que la investigación relacionada con el sensemaking es presentada a menudo como un programa de investigación interdisciplinario.

Se puede definir sensemaking como “el proceso a través del cual las personas trabajan para entender asuntos o eventos que son nuevos, ambiguos o que violan de alguna manera las expectativas” (Maitlis & Christianson, 2014).

Existen básicamente dos enfoques en sensemaking: el organizacional, que ha sido ampliamente estudiado (por una revisión detallada ver (Maitlis & Christianson, 2014) y el psicológico.

Desde un punto de vista psicológico, Dervin define *sensemaking* como relacionado con los comportamientos del individuo, ya sean cognitivos o procedimentales (acciones realizadas) que le permiten construir su movimiento a través del espacio/tiempo, por lo que para ella el *sensemaking* es un comportamiento relacionado con la comunicación (Dervin, 1983).

Por otra parte, la visión organizacional sostiene que cuando los miembros de una organización encuentran momentos de incertidumbre o ambigüedad buscan aclarar qué está sucediendo por medio de la extracción e interpretación de pistas en sus entornos usándolas como base para una interpretación plausible que provea de orden y “tenga sentido” (*makes sense*) sobre lo que ha ocurrido (Maitlis & Christianson, 2014).

El *sensemaking* incluye la tarea de interpretación aunque es más que ella, ya que puede incorporar cierto nivel de invención: “interpretación and *sensemaking* son bastante diferentes, *sensemaking* implica autoría así como interpretación, creación así como descubrimiento” (Weick, 1995).

Una idea básica en *sensemaking* es que la realidad es un logro que se alcanza continuamente y que emerge de los esfuerzos de crear orden y darle un sentido retrospectivo a

lo que sucede (Weick, 1993). En el modelo propuesto en esta Tesis, también la realidad (el sentido de lo expresado) es un producto que emerge de la colaboración de los agentes los cuales consultan a sus pares al encontrar mensajes con vaguedad o incertidumbre, buscando mejorar la calidad de la respuesta que dan.

Obsérvese que si bien el *sensemaking* fue definido para organizaciones sociales de personas, sus ideas básicas pueden aplicarse a organizaciones que incluyan agentes colaborativos cualesquiera (por ejemplo, dos sistemas inteligentes que se comunican). Desde el punto de vista práctico, *sensemaking* es un proceso que busca la creación de un entendimiento de la situación en situaciones de alta complejidad o incertidumbre de forma de tomar decisiones.

2.7. *Soft sensors*

Un *soft sensor* (SS) puede ser definido como “la asociación de un sensor *hardware* que permite la medida *on-line* de algunas variables de proceso con un algoritmo (*software*) con el objetivo de proveer estimaciones *on-line* de las variables no medidas” (De Assis & Filho, 2000). De forma similar, para Napoli y Xibilia (Napoli & Xibilia, 2011) son “sistemas consistentes de algoritmos matemáticos que producen estimaciones confiables en tiempo real de variables no medidas usando su correlación con los datos disponibles”. Dos características básicas son propias de estos sensores. Por una parte, pueden compartir información que contenga vaguedad, como en el caso de (Gadeo-Martos *et al.*, 2011) en la cuales los sensores utilizan un motor de razonamiento *fuzzy*, y por otra, tener un cierto grado de error, ya sea de medición (para los *hard sensors*) como de predicción (en el caso de los SS).

Otra faceta de los SS es el mantenimiento que implica un cambio en la integración del mismo. Dicho mantenimiento es uno de los principales problemas que tienen los SS y se debe realizar debido a que los sensores componentes envejecen (por desgaste mecánico, corrosión, etc.), el entorno cambia e incluso los mismos procesos en los que se usan cambian. Por ello se suele buscar diseñar SS adaptativos, usando técnicas tales como el *sliding window*, *ensemble methods* y *just-in-time learning* (Saporo, 2014).

Los sensores *hardware* (o simplemente sensores) son equipos físicos destinados a

obtener ciertas observaciones (mediciones) sobre un fenómeno físico de interés, por ejemplo, la temperatura ambiente de una habitación. Se definirán los sensores virtuales como una generalización de los anteriores, como un software que obtendrá información sobre fenómenos que acontecen en su entorno virtual, típicamente una computadora o una red de ellas, por ejemplo, el *software* que detecta que un programa está intentando realizar cierta operación en la red. Obsérvese que si bien un sensor virtual es *software* no lo llamamos *Soft Sensor* (SS) ya que las entidades (también virtuales) que observa y mide están en el mismo plano ontológico que él. Un *soft sensor* podría ser además virtual, si las mediciones que utiliza son hechas por sensores virtuales.

Si se consideran las definiciones hechas en el apartado 2.5 un *soft sensor* será siempre calculado por algún agente colaborativo perteneciente a algún espacio inteligente generalizado.

2.7.1. *Soft sensors* distribuidos

La investigación sobre sistemas distribuidos de SS ha sido abundante, tal vez debido a que los sistemas con muchos sensores tienen la ventaja de mayor rendimiento, confiabilidad y facilidad de supervisión que los de sensores únicos (Aziz, 2014). Los sistemas con inteligencia distribuida han probado tener capacidad de evaluar autónomamente su entorno, aprender de sus experiencias y comunicar su conocimiento conclusiones y razonamiento a los otros. Esta técnica es llamada “redes cognitivas de sensores” (Soares *et al.*, 2011).

Aziz *et al.* (Aziz, 2014) utilizan una serie de sensores cuyas mediciones se combinan usando fusión de información en un nodo centralizador. En (Aziz, 2008) se propone un enfoque multinivel para sistemas de sensores geográficamente distribuidos donde el centro de fusión combina las decisiones *soft* de los sensores, pero no se toma en cuenta la confiabilidad de los mismos. La confiabilidad en los sensores sí es tomada en cuenta en (Aziz, 2011). La confiabilidad en un sensor es considerada en el modelo propuesto en esta Tesis y es usada continuamente para elegir los peers con los que interactúa cada elemento.

En cuanto a su rendimiento en sistemas multinivel se ha estudiado para redes de sen-

sores *wireless* (WSN) (Niu *et al.*, 2006; Soares *et al.*, 2011; Cheng & Wang, 2012). Una WSN es un sistema distribuido que es compuesto de unidades autónomas con capacidades de detección (*sensing*) interconectadas por sistemas de comunicación inalámbricos. Las WSN ofrecen soluciones potencialmente de bajo costo a problemas tales como rastreo de los objetivos militares, monitorización de la salud, sistemas de control de entorno, monitorización de animales, casas inteligentes, etc. En esos trabajos se compara el rendimiento centralizada frente a la descentralizada, pero no se analiza la misma en el caso de que no haya un elemento central en la red (como sucede en una red P2P). La idea de que un sensor colabore no sólo con algunos vecinos sino con todos los sensores del sistema ya aparece en (Gadeo-Martos *et al.*, 2011).

2.7.2. Sistemas de *soft sensors* basados en la naturaleza

Varios autores han propuesto sistemas de SS inspirados en la naturaleza, siendo los enfoques más comunes las redes neuronales y los enfoques evolutivos o híbridos.

Sobre el uso de Redes Neuronales Artificiales (RNA, o ANN, *Artificial Neural Networks*, en inglés) hay numerosos trabajos al respecto, por ejemplo (Bo *et al.*, 2003; Devogalaere *et al.*, 2002; Fortuna *et al.*, 2005; Pani *et al.*, 2016; Rani *et al.*, 2013). Las ventajas que tienen las RNA respecto a un modelo construido a partir de las leyes físicas y las relaciones que gobiernan el sistema son su sencillez y que no se precisa tener conocimiento del sistema gobernado para modelar el SS. Otra de las ventajas que las hace atractivas es que pueden aproximar cualquier función continua Entrada→Salida (más precisamente, una red *feed-forward* con una sola capa oculta que contiene un número finito de neuronas, por ejemplo, un perceptrón multicapa, puede aproximar funciones continuas en subconjuntos compactos de $\mathbb{R}^n$ con tal que la función de activación cumpla una serie de requisitos poco estrictos) (Cybenko, 1989; Hornik, 1991). Dentro de las topologías más usadas se encuentran los perceptrones multicapa (MLP, *MultiLayer Perceptron*, en inglés (Hornik, 1991; Gonzaga *et al.*, 2009), las funciones de base radial (RBFs) (Napoli & Xibilia, 2011) y las *Self-Organizing Maps* (SOMs) (Alhoniemi *et al.*, 1999). Entre sus principales desventajas está la difícil interpretación del conocimiento aprendido por la red. A los efectos de poder modelar sistemas en los que las mediciones resultan en valores lingüísticos (“alto”,

“turbio”, etc.) algunos las usan formando sistemas *neuro-fuzzy* (Bouhouche *et al.*, 2008; Mozaffari & Azad, 2015) citándose como modelo de SS *neuro-fuzzy* a la red ANFIS (sistema de inferencia daptativo neuro-fuzzy o *Adaptive Neuro-Fuzzy Inference System*) (Azadeh *et al.*, 2011). Las RNA también se combinan con *Partial Least Squares* y con *Principal Component Analysis* (PCA) (Kadlec & Gabrys, 2007; Kordon *et al.*, 2002). En (Pu *et al.*, 2011) se integran las RNA con algoritmos genéticos y en (Wang & Han, 2015) con tecnologías *swarm*. Finalmente, la capacidad de generalización de las RNAs se puede aumentar (es decir, mejorar el resultado obtenido) usando técnicas de *bootstrap* o por *stacking* (Napoli & Xibilia, 2011).

Uno de los casos de estudio de esta Tesis es un *soft sensor*, tal como es explicado en el Cap. 6. Dado que el modelo propuesto sugiere usar redes neuronales, se puede decir que el *soft sensor* así construido está inspirado en la naturaleza.

2.8. Fundamentos de las estructuras holónicas

2.8.1. Holones

2.8.1.1. Definición y generalidades

En 1926 Smuts (Smuts, 1926), siguiendo a Aristóteles, sostuvo que el todo es más que la suma de las partes e introdujo el concepto de holismo, apuntando que, a través de los “todos”, se crean la variedad y complejidad en el universo. Smuts definió el holismo como “la tendencia de la naturaleza de formar todos que son mayores que la suma de las partes, a través de la evolución creativa”.

Posteriormente, continuando con el estudio de los “todos”, el filósofo austríaco Arthur Koestler acuñó el término holón. En su libro “The ghost in the machine” (1967) (Koestler, 1982) comenta que las jerarquías son algo fundamental en la naturaleza: al tener una organización jerárquica, un sistema se vuelve más resistente, fácil de mantener, de regular y de reparar que uno desorganizado. Anota, además, que la principal característica de las jerarquías es su relatividad: “todo” y “parte” en el sentido absoluto no existen en ninguna parte (Koestler, 1982) sino que lo que sucede realmente es que las entidades presentan dos

caras: una mostrando un todo auto-contenido y otra mostrando cada parte dependiente. A esta propiedad de poder ser considerado todo o parte a la vez se la suele conocer como efecto Jano, por las dos caras: una visión como una totalidad y otra como una parte: “El holón es como un Jano de dos cabezas: observa su propio interior, es decir, los niveles más bajos de la estructura jerárquica, ya que se considera un todo formado por partes subordinadas, y observa su exterior cuando se considera una parte de un todo más vasto. Sin embargo, si se observa a si mismo se ve como un individuo único e independiente que debe sobrevivir” (Mella & Gazzola, 2017).

Todas las entidades son pues, para Koestler, un “todo” y una “parte” a la vez, y para nombrarlas, el término propuesto fue holón, del griego *holos* (todo o totalidad) más el sufijo *on* como en protón o neutrón, sugiriendo una partícula o parte. Dichas entidades pueden ser seres vivientes u organizaciones sociales (Koestler, 1982).

El concepto de holón no es el de un objeto o entidad específica sino que representa un enfoque lógico para observar objetos o entidades basado en la relación todo-parte; el “holón” es una idea, la base de una visión del mundo: la visión holónica (Mella & Gazzola, 2017).

Este enfoque holónico se ha usado en una cantidad de disciplinas y contextos: física (Capra, 1996, 1983), ingeniería (Samir Dani, 2004; Botti & Giret, 2008; Mc.Hugh & Merli, 1995; Soriano, 2003), robótica (Iwamura *et al.*, 2006), control (Hall *et al.*, 2005), biología (Shafaei & Aghaei, 2008), estudios organizacionales (Fischer *et al.*, 2003), gestión (Chirn & McFarlane, 2005)(Zhang *et al.*, 2003) y sistemas de producción y de cadena de suministros (Amiri, 2006; McFarlane & Bussmann, 2000) entre otros.

En la naturaleza los sistemas orgánicos tienen vitalidad, estructura y propósito; son capaces de adaptarse y de reaccionar a estímulos externos, lo que se logra gracias a su estructura holónica. En esos sistemas hay un aspecto ontogénico en el que cada holón tiene alguna excelencia en la capacidad de desarrollar los procesos, y sobrevive gracias a ella. Hay también un aspecto filogénico: el sistema holónico sobrevive porque la combinación de las capacidades, competencias o habilidades centrales (*core*) individuales de los holones es más poderosa y flexible que las de los miembros por separado (Mc.Hugh & Merli, 1995; Marik *et al.*, 2007, 2013). Ejemplos de holones naturales son una célula, un hígado humano o el sistema solar; un sistema multi-agente autónomo holónico o una

Disciplina	Primer nivel	Segundo nivel	Tercer nivel
Física	Partículas	Átomos	Moléculas
Química	Moléculas	Compuestos	Bases
Genética	Bases	DNA	Genes
Gobierno	Comunidades/pueblos	Regiones/estados	Naciones
Ingeniería	Componentes	Mecanismos	Máquinas

Tabla 1: Ejemplos de holarquias según Turnbull

fábrica organizada holónicamente podrían ser ejemplos de holones artificiales. Otros ejemplos se dan en la tabla 1 (tomado de (Turnbull, 2001)). Así en la química el nivel más elemental que interesa es la molécula, a partir de las cuales se forman los compuestos, y con los compuestos, por ejemplo las bases o los ácidos, similarmente en ingeniería las piezas componentes (suponiendo que tienen algún grado de autonomía) se reúnen para formar mecanismos (también con autonomía) que formaran finalmente máquinas.

Los holones pueden ser vistos como agentes abstractos con un cuerpo (es decir, agentes que tienen una parte física que interactúa con su entorno, y que pueden estar formados por otros agentes abstractos) (Benaskeur *et al.*, 2007).

Para Koestler (Koestler, 1982) un holón es una unidad que es:

- autónoma, con una función y dinámica que es distinta del contexto.
- viable porque tiene una forma estable que le permite lidiar con los disturbios del entorno para sobrevivir.
- independiente (que confía en sí misma).
- dependiente porque está sujeta a alguna forma de control por la unidad superior ya que tiene un rol para la supervivencia de la estructura en la que está contenida.
- interactiva, o sea, verticalmente vinculada a las unidades superiores e inferiores.
- caracterizada por reglas de comportamiento que definen las restricciones al actuar del holón tanto al considerarlo como todo y también al considerarlo como parte.

Mella y Gazzola (Mella & Gazzola, 2017) generalizan la noción de holón a cualquier objeto o concepto observable en tres niveles:

- 1) como una unidad autónoma e independiente que actúa de acuerdo a sus propios cánones de comportamiento,
- 2) como una unidad de nivel superior poseyendo propiedades emergentes con respecto a las partes componentes a las que trasciende, y
- 3) como una unidad subordinada, ya que es parte de un todo más vasto que la condiciona.

Una jerarquía, en la naturaleza, es un orden creciente de “totalidad” (de partículas a átomos a células etc.) y debería ser llamada una holarquía (Wilber, 2000).

Las redes holónicas son, al igual que los holones, entidades conceptuales cuyos nodos son holones.

El holismo es típicamente evolucionista y se concentra en la globalidad del todo más que en la parte, considerando al todo como caracterizado por propiedades emergentes que no son encontradas en los componentes (lo que queda claro por la propiedad b) anterior). Este enfoque contrasta con el reduccionismo: mientras que el holismo sostiene que el todo es más que la suma de las partes, el reduccionismo plantea que lo simple es la fuente de lo complejo y que para explicar los fenómenos complejos es necesario “reducirlos” a sus elementos constitutivos (Mella & Gazzola, 2017).

Como puede verse de la idea de todo y parte, la visión holónica considera relevante no tanto la conexión entre elementos sino su inclusión en otros. Cada holón incluye a aquellos de un nivel inferior pero no puede ser reducido a ellos, él los trasciende a la misma vez que los incluye, presentando entonces propiedades emergentes (Edwards, 2018; Simon, 1996). Por ello, la holarquía puede ser vista como un sistema multi-capas o multi-nivel con una estructura de árbol (Pichler, 2000).

En términos generales, los holones cooperan en una forma benevolente para alcanzar una meta; no es una benevolencia ciega donde cada holón pasa todo su tiempo en cooperaciones con los otros; al contrario, él considera constantemente sus compromisos y deberes y rechaza colaborar solo cuando las acciones solicitadas son imposibles o al-

tamente desfavorables al proceso, y, en este sentido, los holones pueden ser llamados semi-autónomos.

Desde un punto de vista de la ingeniería, como en el caso de sistemas multi-agente, existen dos formas de cooperación: explícita (el compromiso es establecido a través de una comunicación) e implícita (los holones son diseñados de manera que un comportamiento orientado hacia una meta emerja a partir de los comportamientos de los sub-holones, considerados como agentes).

Los holones han sido relacionados con los agentes abstractos (Botti & Giret, 2008), es decir, agentes formados por otros agentes: en ambos casos hay una auto-similaridad, un estar compuesto por elementos del mismo tipo.

El enfoque holónico permite tener estructuras que evolucionan con el tiempo basadas en reglas simples y locales, y que no se apartan mucho de su punto de equilibrio (Bourjot *et al.*, 2005).

La visión holística es típica del pensamiento sistémico (Mella, 2012), según la cual la unidad básica de observación es el sistema, donde cada sistema debe ser observado verticalmente en sus interacciones con el super-sistema y los subsistemas componentes y horizontalmente en sus interacciones con otros sistemas paralelos con el objetivo formar una red cuyos límites son dados por la necesidad de observación más que por los límites intrínsecos del sistema.

Koestler distingue entre holarquías de salida o descendentes, donde el holón de nivel superior solo coordina actividades y manda información a los holones de menos nivel que son los que realizan las tareas, y holarquías de entrada o ascendentes, en las que en cada nivel los holones filtran y sintetizan las entradas que les llegan de sus holones subordinados. En especial, Shimizu (Shimizu, 1987) define un computador cognitivo autónomo (*autonomic cognitive computer*) como una holarquía de entrada a un conjunto de holones que procesan información en paralelo con el objetivo de producir, por síntesis, información en el nivel final, cada nivel sintetizando información del nivel inferior.

Según Mella y Gazzola: “una organización puede ser entonces considerada como un macrosistema formado para alcanzar un macroobjetivo. Por lo tanto podemos compararla con un computador autónomo cognitivo (*autonomic cognitive computer*), es decir, una holarquía formada por órganos anidados, cada uno de los cuales está especializado en

alcanzar una parte del macroobjetivo” Mella & Gazzola (2017).

Mella y Gazzola (Mella & Gazzola, 2017) distinguen tres tipos de holarquías:

- 1) modulares o estructurales, en las que los holones se agrupan de acuerdo a características estructurales o cualitativas (como en los sistemas de clasificación, donde cada clase tiene propiedades cada vez más específicas).
- 2) cognitivas auto-organizadas, formadas por individuos con sensores interconectados por relaciones o por programación, coordinación y control. Los sistemas abiertos jerárquicos de Koestler y el kosmos de Wilber son de este tipo.
- 3) operativas, que son un arreglo de holones que se conectan para ir formando estructuras operativas cada vez más grandes. Los holones son vistos en ese caso como agentes-operadores elementales.

Una de las principales características de los holones es su granularidad múltiple que se manifiesta a través de la replicación en estructuras auto-similares de tipo fractal con múltiples niveles de resolución: la holarquía.

Hock (Hock, 1994) llama a la holarquía “*caord*” de las palabras caos (*chaos*) y orden (*order*): “cualquier comunidad no lineal, adaptativa, compleja y auto-organizada, sea física, biológica o social”. La holarquía es entonces cualquier conjunto de holones “caóticamente ordenado”.

El concepto de holón ha recibido además otros nombres de parte de distintos investigadores; por ejemplo Simon los llama “sub ensambles” (“sub assemblies”) (Simon, 1957) y formas intermedias (“*intermediate forms*”); Ingsber los llama estructuras estables y Hatvany entidades (citado por (Turnbull, 2003)).

El paradigma holístico ha llevado a la formulación de cosmologías completas tales como las de Wilber (Wilber, 2007, 2000) y de Laureyssens ((Laureyssens, 2009).

La holarquía cognitiva de Koestler es el punto de partida de Wilber. Wilber trata de generalizar la noción de holón: “el mundo no está compuesto de átomos, símbolos o células o conceptos. Está compuesto de holones” y propone una cosmología (en el sentido pitagórico de “*kosmos*” como la unión de todos los dominios de existencia, desde la material al espíritu, no solo en el mundo físico) en la cual los holones juegan un rol

fundamental: la realidad está formada por todos/partes, es decir, los holones (este es el primer principio de la cosmología de Wilber) (Wilber, 2000). Wilber tiene una visión metafísica de la evolución hacia la conciencia y concibe al *kosmos* como una holarquía cognitiva general que produce la evolución de la naturaleza hacia la conciencia, es decir, entiende los holones como unidades de conciencia.

El segundo principio es: cada holón tiene cuatro fuerzas que lo dirigen permanentemente: agencia, comunión trascendencia y disolución. Las capacidades que permiten al holón ser parte/todo son llamadas capacidades horizontales, las capacidades verticales son las relacionadas con la disolución (tal como cuando una célula se disuelve en moléculas luego de su muerte) y las de trascendencia, cuando el holón se vuelve parte de algo más complejo (por ejemplo cuando un conjunto de moléculas se vuelven una célula). El conflicto constante entre las necesidades de ser una parte de forma de subsistir y la autonomía implican una tensión real (Boardman & Sauser, 2008) que implica la búsqueda de un balance a través de cualquiera de los comportamientos del holón. Su tercer principio se relaciona con la emergencia: los holones emergen. Un holón no puede ser reducido a sus componentes en una forma creativa. Cada nivel de complejidad debería ser acompañado por un cierto grado de conciencia (Wilber, 2007) (por ejemplo, para un reptil, la conciencia está formada por hambre y otras sensaciones fisiológicas y sensoriales). El resto de los principios no están relacionados directamente con el concepto de holón.

Laureyssens da una interpretación física (topológica) al concepto de holón y considera a los holones como regiones discretas del espacio tales que una región penetra a otras generando los holones (Laureyssens, 2009).

Por último, en varios campos de la física el término holón es usado como sinónimo de *chargon*. En este caso, una excitación electrónica puede ser descompuesta en dos excitaciones independientes, una contribuyendo con el *spin* (el *spinon*) y la otra la carga (el holón) (Auslaender *et al.*, 2005; Bernevig *et al.*, 2001; Koitzsch *et al.*, 2006). El significado dado al término holón en esta Tesis coincide con el dado por Koestler y Wilber, que por otra parte es el que ha adoptado la industria como lo muestran los ejemplos previos.

“La nueva perspectiva holónica nos fuerza a observar el universo circundante (a un nivel físico y biológico, y en el sentido real o formal) a tomar en cuenta el todo/parte

(holarquías) y las relaciones de antes/después (redes de holones)” (Mella & Gazzola, 2017).

2.8.1.2. Algunos sistemas holónicos notables

Algunas holarquías han tomado especial notoriedad, debido principalmente a su impacto en la industria:

- 1) Sistemas holónicos de manufactura (*Holonic manufacturing systems*): involucran unidades simples que son holones y métodos para la cooperación para realizar tareas complejas. Si una unidad falla o tiene problemas, las otras unidades cambian la secuencia para evitar problemas de forma de manejar los problemas bottom-up. El comportamiento del sistema emerge de reglas simples y es muy sólido y adaptable.
- 2) Sistemas biónicos de manufactura (*Bionic manufacturing systems*): es caso especial de holarquía de producción donde los operadores son considerados holones, que se agrupan en módulos y dispuestos en varios niveles para formar una holarquía, como si fuera un sistema biológico. Lo que caracteriza a estos sistemas es que las unidades operativas o su agrupamiento pueden decidir autónomamente si realizar un proceso y qué entradas (y salidas) necesitan (y generan) a partir de la información de qué parte del todo debe ser completada por el módulo.
- 3) Sistemas fractales de manufactura (*Fractal manufacturing systems*): es una holarquía de abajo-arriba (*bottom-up*) formada por pequeños componentes que son módulos autónomos cuya lógica operativa se repite en varios niveles verticales reproduciendo en cada nivel las características de la estructura completa.

Uno de los objetivos de las organizaciones holónicas es crear o implementar sistemas de manufactura ágiles es decir, sistemas altamente flexibles y automatizados, de allí su interés en la industria. Una red de empresas es una red holónica formada de firmas autónomas que están ubicadas en distintas partes, caracterizadas por distintos roles y operaciones pero integradas en términos de misión visión y la meta de su negocio común y conectadas por una red holónica, real o virtual, orientada a menudo a alcanzar algún objetivo común

Figura 1: Esquema cabeza-cuerpo del holón

por medio de compartir información, recursos y competencias sin ninguna restricción jerárquica de subordinación. Las redes de empresas son llamadas firmas holónicas o firmas virtuales.

Cuando la relación entre los holones se alcanza usando una red de información la red se vuelve una verdadera organización virtual (o sea cuando las relaciones a través de la red permiten el fluir de insumos y productos -ambos serían datos - para sus integrantes)

2.8.1.3. Estructura.

La estructura de holón que se considerará en este trabajo de Tesis es la de (Fischer *et al.*, 2003) de cabeza-cuerpo, donde la cabeza realiza toda la comunicación hacia afuera de holón y coordina las actividades del cuerpo; se puede decir que la cabeza representa al holón hacia afuera. El cuerpo realiza las demás funciones del holón, tal y como muestra la figura 1 (las flechas representan comunicaciones y H1, H2... los holones). Ejemplos de autores que han utilizado esta arquitectura son (Fischer *et al.*, 2003; Schuldt, 2011; Beheshti *et al.*, 2016a; Schillo & Fischer, 2004; Unland & Ulieru, 2005).

2.8.1.4. Propiedades de los holones

La propiedad clave de los agentes holónicos es su autonomía limitada. Los agentes en un sistema holónico son racionales y auto-interesados y deciden qué hacer en base a sus

propias preferencias como hacen los agentes en los sistemas multi-agente (MAS por su sigla en inglés) comunes y corrientes. Pero al unirse un agente a un holón no puede hacer lo que quiera, debe cumplir con los compromisos del holón. Esto no significa que cada acción que el agente tome esté determinada por la cabeza del holón como sucedería en un MAS totalmente cooperativo. Un agente usualmente puede decidir cuándo unirse a un holón o dejarlo (Beheshti *et al.*, 2016b).

Turnbull (Turnbull, 2000) citando a Mathews (Mathews, 1996) distingue entre holones fuertes y holones débiles (*weak holons*). Los fuertes son los holones que pueden existir autónomamente mientras que los débiles son aquellos que no pueden existir sin el resto de los miembros de la holarquía. En este trabajo no se distinguirá entre ellos ya que no interesa estudiar su supervivencia sino su formación.

Las holarquías son jerarquías dinámicas que se caracterizan por hacer uso eficiente de sus recursos, ser robustas frente a cambios en su entorno y ser flexibles y adaptables, formadas por holones (Botti & Giret, 2008). Se puede decir que las holarquías garantizan la estabilidad, predecibilidad y optimización global del control jerárquico (Valencia-Jiménez & Fernández-Caballero, 2008). Esas holarquías no son ni una jerarquía estática ni una estructura centralizada, aunque tienen propiedades de ambas, y se corresponden con las estructuras *edge* definidas por Alberts y Hayes (Alberts & Hayes, 2005). Las estructuras *edge* son básicamente holarquías, en la tabla 2 (basada en una similar en (Alberts & Hayes, 2005) puede verse la relación entre jerarquías, estructuras *edge* y holarquías.

El control se distribuye de forma que las actividades asociadas a él no recaen todas en el holón que delega sino que se son parcialmente cumplidas por el holón que recibe lo delegado. Esto a su vez permite que las comunicaciones sean menos (Schwaiger & Stahmer, 2005).

Cuando la racionalidad es limitada (capacidad de procesamiento de información y de comunicaciones) y se desea tener una buena velocidad de adaptación al entorno, los sistemas formados por bloques 'grandes' tienen mejor desempeño que los formados por bloques 'pequeños' (Valckenaers & Van Brussel, 2005), de ahí la importancia de los holones (entendiendo por bloque a un conjunto de agentes que actúan coordinadamente). Simon demuestra esto con su parábola del relojero (Simon, 1996).

	<i>Jerarquías</i>	<i>Organizaciones edge</i>	<i>holarquías</i>
<i>Mando (Comando)</i>	Por directiva	Estableciendo condiciones	Estableciendo condiciones (metas y restricciones)
<i>Liderazgo</i>	Por posición	Por competencia	Por competencia
<i>Control</i>	Por dirección	Una propiedad emergente	Una propiedad emergente
<i>Toma de decisiones</i>	Función de la línea de mando	El trabajo de cada uno	El trabajo de cada uno
<i>Información</i>	Acumulada	Compartida	Algo se puede acumular localmente y algo se puede compartir
<i>Flujos de información predominantes</i>	Vertical, acoplado con la cadena de mando	Horizontal, independiente de la cadena de mando	Horizontal, independiente de la cadena de mando, respetando el nivel del holón
<i>Gestión de la información</i>	<i>Push</i>	<i>Post-Pull</i>	Principalmente <i>Post-Pull</i>
<i>Fuentes de información</i>	Monopolios, arriba-abajo	Mercados eclécticos y adaptables	Mercados eclécticos y adaptables
<i>Procesos organizativos</i>	Prescriptos y secuenciales	Dinámicos y concurrentes	Dinámicos y concurrentes
<i>Individuos ejecutores</i>	Restringidos	“empoderados”	“empoderados”

Tabla 2: Relación entre jerarquías, estructuras edge y holarquías

2.8.1.5. Formación de los holones y holarquías

Pocos han sido los intentos por formalizar la generación de holones dentro de un sistema. Ulieru (Ulieru, 2002b,a) demuestra que cuando la holarquía se asocia a la realización de un plan que puede dividirse en sub planes cuyas definiciones pueden tener tareas en común (es decir cuya definición es vaga), la holarquía minimiza la entropía *fuzzy* del conjunto de agentes. Un enfoque basado en la utilidad (medida de cualquier forma deseada, no solo económica: estatus, seguridad, etc.) que obtiene el holón por formar parte de la holarquía es expuesto en (Paggi Straneo *et al.*, 2010). Allí se afirma que la condición suficiente para que una serie de holones se integren para formar una holarquía es que la mínima utilidad esperada por alguno de ellos al pertenecer a la holarquía sea mayor que la máxima esperada por ese holón si no pertenece a ella (es decir, el holón obtiene algún tipo de beneficio por pertenecer a la holarquía) mientras que es necesario que la mínima utilidad individual de cualquiera de los holones sea mayor que la mínima que obtendría al pertenecer a ella (o sea, que no llegue a perder utilidades por pertenecer al grupo). Teniendo en cuenta la incertidumbre, las dos condiciones anteriores serían:

- que es necesario que la mínima utilidad esperada sea mayor que la mínima individual posible y,
 - que es suficiente si la mínima esperada es mayor que la máxima individual posible.
- Estas condiciones pueden ser vistas como una generalización al criterio de *core* no vacío en un juego cooperativo (Andrade *et al.*, 2015; Chalkiadakis *et al.*, 2011) (Curiel, 1997): el holón surge cuando el conjunto de sus agentes tienen un *core* no vacío al considerar su actuar como un juego cooperativo.

Este enfoque de juego es muy adecuado al estudio de la auto-organización y las posibles asociaciones “naturales” que se generen entre los agentes ya que estudia la toma de decisiones estratégicas en escenarios donde puede haber acuerdos para asociarse y donde los agentes actúan, por tanto, colectivamente (Andrade *et al.*, 2015) aunque pueda existir competencia: los agentes cooperan para alcanzar una meta global pero compiten por las capacidades de comunicación y otros recursos. Andrade *et al.* (Andrade *et al.*, 2015)

utilizan este enfoque para la elección de *partners* en una holarquía. A su vez, de los juegos cooperativos, los que más interesaran aquí serán los de función característica por lo siguiente: un juego de ese tipo está dado por un par (N, v) donde N es el conjunto de jugadores en el juego, y $v : 2N \rightarrow \mathbb{R}$ es una función que da valor real $v(C)$ a cada conjunto de jugadores C contenido en N . La función v es la función característica. $v(C)$, es el monto que la coalición C podría ganar si ellos deciden cooperar. Dado que en esta Tesis se consideran las estructuras que forman los individuos (típicamente, agentes), este enfoque es muy pertinente: los agentes cooperan para alcanzar la meta global, pero donde a la vez se compite por las capacidades de comunicación.

Una comparación entre el trabajo de Andrade et al. (Andrade *et al.*, 2015) y el modelo propuesto en esta Tesis puede verse a continuación:

- Elección de *partner*:

Andrade et al.: Se enfoca en cómo elegir el partner para formar la empresa virtual (VE). Una de las características de la VE es tomar la responsabilidad de la meta global. La VE está formada por member enterprises (ME) que pueden ser toda una holarquía (por ejemplo una VE).

En esta Tesis: Se elige el partner con quien colaborar de forma de tener la mayor calidad posible en el resultado. Por posible se entiende: se busca la respuesta entre quienes han tenido un mejor desempeño respecto al agente consultante (digamos, una medición totalmente local de la calidad) y que responden a la consulta. La Virtual Enterprise está formada por todos los agentes que pueden responder al agente externo Ω . La Member Enterprise (ME) se corresponde con subconjuntos de agentes que responden a (predicen) los distintos campos dados una configuración de mensajes disponibles en un instante y a las calidades propias de cada uno. La meta es responder el mensaje, lo que puede corresponder por ejemplo realizar una clasificación de las entidades representadas por los datos del mensaje. La ME realiza un conjunto de las tareas (predecir un conjunto de campos) de forma que se cumplan ciertas condiciones. Están formadas por individuos cooperativos.

- Agentes y éxitos:

Andrade et al.: Los agentes que responden pueden ser vistos como los más exitosos en la colaboraciones previas (para pensarlo como Jarimo y Salo (Jarimo & Salo, 2009).

Ellos proponen incluir las colaboraciones exitosas previas en la evaluación de la utilidad de incluir un holón para formar la holarquía).

En esta Tesis: Los agentes no responden si no tienen una mejor respuesta (mayor calidad). Ver las consideraciones sobre un umbral de aceptación y el concepto de cohesión en (Paggi Straneo *et al.*, 2010).

- Metas y habilidades:

Andrade et al.: Las metas se asocian a habilidades (*skills*) y obligaciones. Una obligación tiene asociada una o más habilidades y un rol una o más obligaciones.

En esta Tesis: Una habilidad (*skill*) de un conjunto de agentes es la aptitud predecir con buena calidad un cierto campo y la habilidad de optimizar sus mensajes. El rol sería poder predecir cierto campo

- Roles:

Andrade et al.: Un rol está determinado por un conjunto de obligaciones.

En esta Tesis: Los roles están asociados a la predicción de cada campo (tareas). Surgen por evolución: en principio cada agente puede predecir todos los campos compuestos. En el presente modelo cada agente puede jugar varios roles (predecir varios campos) y puede ser consultado por ellos.

- Habilidades

Andrade et al.: Las habilidades tienen distinta importancia en la formación del holón.

En esta Tesis: Los pesos mencionados para las skills se corresponden con un vector de pesos W que maneja el sistema, que indica la importancia relativa de las calidades de cada campo en la respuesta final.

- Mejor holón:

Andrade et al.: El mejor holón es el que está en el tope del ranking al ordenar según la función v .

En esta Tesis: El mejor holón es el que predice con mayor calidad (o sea, el que está en el tope del ranking si se lo ordena por número de veces que dio la mejor respuesta).

- Función característica:

Andrade et al.: Es v = valor creado cuando los miembros de S (agentes) se agrupan e interactúan sin la cooperación de otros jugadores. Se busca la colaboración que maximice v .

En esta Tesis: Cada jugador es un agente, v , que es la calidad en la respuesta y que básicamente será inversamente proporcional al error cometido al predecir, puede tener distintas expresiones analíticas. De esa manera lo que se logra es dar una mejor respuesta dado el conjunto de agentes que se pueden acceder (es decir, que responden si se les consulta) desde el que plantea la consulta (llamémoslo Y) en un cierto momento en un solo paso. Sea $F_Y(t)$ ese conjunto. Entonces lo que se devuelve es $\underset{F_Y(t)}{\operatorname{argmax}}(\text{calidad})$ y $\underset{F_Y(t)}{\operatorname{max}}(\text{calidad})$ donde se entiende que argmax es el jugador (agente) que dió mejor calidad.

La función v se define recursivamente como (haciendo $F_Y(t) \equiv \Phi$ para simplificar la notación):

$$\begin{cases} v(\Phi) = g(\%Error_Y) \text{ si } \Phi = \{Y\} \\ v(\Phi) = \max(v(\Phi), v(\Phi|Y)) \text{ si } \Phi \supset \{Y\} \end{cases} \quad (Ec.5)$$

donde $\%Error_Y$ denota el error de predicción producido por el agente Y y g es una cierta función monótona decreciente respecto a $\%Error_Y$, por ejemplo:

$$g = \frac{1}{1 + \%Error_Y} \quad (Ec.6)$$

Nótese que v así definida es en realidad dependiente del tiempo:

$$v = v(t)$$

Se está suponiendo que el tiempo que lleva calcular la predicción, etc. es despreciable frente al de la comunicación y que el instante t mencionado es el del comienzo de la consulta al grupo de agentes

Téngase en cuenta que:

- La fusión de información es intrínsecamente jerárquica, ya que de informaciones más elementales se obtiene una de mayor jerarquía.
- En esta Tesis el cálculo de v y la optimización que se hace es local ya que depende del agente inicial, el que empieza la cadena de consulta y de la configuración de mensajes disponibles la que a su vez se ve afectada por los tiempos de respuesta aleatorios de cada agente.

- Considerando el modelo propuesto en el Cap. 5, se está suponiendo aquí que los mejores valores de los subcampos de un campo generan el “mejor” campo (por lo que alcanza con buscar los agentes que mejor predigan los subcampos). Sin embargo, podría ser que con una predicción no tan buena de A , lograrse una mejor predicción de C si está determinado a partir de $A, B, \dots$. Para atenuar esas situaciones y evitar los mínimos locales de v se usa una elección inspirada en el recocido simulado: la probabilidad de que un agente responda cuando su respuesta tiene menos calidad del que lo consultó es $(1 - Q) * MR$ donde MR es el porcentaje de mensajes que le quedan³ y Q su calidad. Por esto, para una Q dada, es menos probable es que responda a medida que se le van agotando los mensajes, es decir, a medida que avanza el proceso de explotación de una solución a la consulta.

Entre los trabajos que con más generalidad estudiaron la formación de holarquías se encuentran el de Fischer, Schillo y Siekmann (Fischer *et al.*, 2003) y el de Gerber, Siekmann y Vierke (Gerber *et al.*, 1999). En particular, en este último se muestra:

- que cualquier sistema multi-agente conteniendo n agentes puede ser mapeado mediante un isomorfismo a un sistema donde solo un agente es representado explícitamente y los demás integrados dentro del entorno el sistema.
- cómo construir un sistema isomorfo de un MAS dado en el cual un grupo de agentes son unidos en un holón.

Gerber prueba además que los holones dentro de un MAS con la operación de composición forman un monoide abeliano (es decir, una estructura algebraica en que la operación es asociativa, conmutativa y existe un neutro - el holón que no hace nada). En las siguientes subsecciones se describe formalmente lo anterior (Gerber *et al.*, 1999).

2.8.1.5.1 Definiciones

Se formalizan a continuación algunos conceptos usados en el teorema de existencia, a saber: sistema multi-agente, función de cambio de estado del sistema y sistemas isomorfos.

³O, en forma más general, de recursos, como energía, etc.

Definición 1. Sistema multiagente .(tomada de (Gerber *et al.*, 1999)

Un sistema multi agente es una tupla $(\mathcal{A}, \varepsilon, \Pi, \Delta)$ donde $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ es el conjunto de todos los agentes. Cada agente α_i es representado por una tupla (S_i, P_i, A_i, ϕ_i) donde

- S_i es el conjunto de todos los posibles estados del agente
- P_i el conjunto de las percepciones posibles
- A_i las acciones que puede realizar
- ϕ_i la función que describe la actividad del agente $\phi_i : S_i \times P_i \rightarrow S_i \times A_i$ Es $\phi(s, p) = (\phi^1(s, p), \phi^2(p, s))$ o sea, ϕ^1 da el nuevo estado del agente y ϕ^2 la acción ejecutada
- ε es el conjunto de los estados posibles del entorno del sistema
- Π es la función de percepción $\Pi : \varepsilon \rightarrow P_1 \times \dots \times P_n$
- Δ es la función de estado del entorno del sistema $\Delta : \varepsilon \times A_1 \times \dots \times A_n \rightarrow \varepsilon$

Definición 2. Función de cambio de estado del sistema

Es la función $\bar{\Delta} : \varepsilon \times S_1 \times \dots \times S_n \rightarrow \varepsilon \times S_1 \times \dots \times S_n$ o sea $\bar{\Delta}(e, s_1, \dots, s_n) = (e', s'_1, \dots, s'_n)$

Definición 3. Sistemas isomorfos

Dos MAS $(\mathcal{A}, \varepsilon, \Pi, \Delta)$ y $(\mathcal{A}', \varepsilon', \Pi', \Delta')$ son isomorfos si existe una biyección $\Psi : \varepsilon \times S_1 \times \dots \times S_n \rightarrow \varepsilon' \times S'_1 \times \dots \times S'_n$ tal que para todo $(e, s_1, \dots, s_n) \in \varepsilon \times S_1 \times \dots \times S_n$ es

$$\bar{\Delta}'(\Psi(e, s_1, \dots, s_n)) = \Psi(\bar{\Delta}(e, s_1, \dots, s_n))$$

2.8.5.1.2 Formación del holón

El teorema que asegura que se puede formar un holón es el siguiente. Obsérvese que una cosa es que se pueda construir un holón y otra es que se produzca (emerja) realmente.

Teorema. Para todo sistema multiagente $(\{\alpha_1, \dots, \alpha_n\}, \varepsilon, \Pi, \Delta)$ y $k \leq n$ se puede construir un sistema multi-agente isomorfo $(\{\alpha', \alpha_{k+1}, \dots, \alpha_n\}, \varepsilon, \Pi', \Delta')$ donde α' es el holon formado a partir de $\{\alpha_1, \dots, \alpha_k\}$.

Este teorema muestra la posibilidad de formación de holones. Más adelante se verá, para el modelo presentado, que la probabilidad de que realmente se forme tiende a 1

rápida­mente cuando la cantidad de agentes en el sistema crece. Gerber et al. (Gerber et al., 1999) prueban la existencia de α' construyéndolo: $\alpha' = (S', P', A', \phi')$ de forma que son, por definición:

- $S' = (S_1 \times \dots \times S_k)$
- $P' = (P_1 \times \dots \times P_k)$
- $A' = (A_1 \times \dots \times A_k)$
- $\phi'(s', p') = ((\phi_1^1(s_1, p_1), \dots, \phi_k^k(s_k, p_k)), (\phi_1^2(s_1, p_1), \dots, \phi_k^2(s_k, p_k)))$
- $\Pi' : \varepsilon \rightarrow (P', P_{k+1}, \dots, P_n)$ tal que $\Pi'(e) = ((\Pi_1(e), \dots, \Pi_k(e)), \Pi_{k+1}(e), \dots, \Pi_n(e))$ para todo e
- $\Delta' : \varepsilon \times A' \times A_{k+1} \times \dots A_n$ tal que $\Delta'(e, a', a_{k+1}, \dots a_n) = \Delta(e, a_1, \dots a_n)$

Una hipótesis que queda implícita para el teorema es que todos los agentes pueden comunicarse y colaborar con todos los demás ya que sino el agrupamiento que se hace para formar el holón podría ser imposible.

En la Sección 7 se muestra la correspondencia entre los conceptos manejados en esos artículos y el presente trabajo, donde queda probado que el modelo propuesto, por su propio diseño, genera estructuras holónicas de agentes.

2.8.2. Informones

“Se ha dicho: el todo es más que la suma de sus partes. Es más correcto decir que el todo es algo más que la suma de sus partes, porque sumar es un procedimiento sin significado, mientras que la relación todo-parte es significativa”. Kurt Koffka (1999). *Principles of Gestalt Psychology*, p.176, Psychology Press

El concepto de informón fue definido por Williams Sulis en el contexto de sistemas dinámicos complejos (Sulis et al., 2003). Sulis plantea que un informón es un aspecto de la realidad que posee la capacidad de informar, sin importar como se interprete esa información: “un informón es un aspecto de la realidad que existe antes que cualquier interpretación (marco semántico)”⁴ (Sulis et al., 2003). En este trabajo se adoptará una

⁴“An informon is an aspect of the reality that exist prior to any interpretation (semantic frame)”

definición similar a la de Sulis, dada por Alonso et al. (Alonso *et al.*, 2004; Martínez Luaces *et al.*, 2008): “informón es el elemento básico de información que tiene sentido para un holón y que le permite tomar las decisiones correctas y ejecutar las acciones apropiadas”. Esas decisiones y acciones generarán nuevos informones que van a afectar al propio holón y a su entorno. Los holones reaccionan a, y planifican en base a, informones.

En el contexto de esta Tesis, los mensajes que intercambian los *peers* son informones que usarán los distintos holones (grupos de *peers*) para tomar sus decisiones.

Los informones pueden ser vistos como perturbaciones desde el entorno a los holones, una serie de estímulos que son recibidos y que afectan los holones y pueden ser considerados como distribuidos en capas semánticas que corresponden aproximadamente a los niveles de los holones afectados. Por ejemplo, una noticia o novedad (tal como el nacimiento de un niño) va a afectar significativamente el comportamiento del padre del bebé (un holón humano), en menor grado el funcionamiento del entorno de trabajo del padre (un holón de mayor nivel) pero seguramente no va a afectar el comportamiento de toda una nación o continente: el informón es significativo (o, para ser preciso, existe) solo para un cierto conjunto de holones agrupados en niveles (ver figura 2). En dicha figura se observa cómo los holones H1, H2 y H3 de un cierto nivel $k + 1$ se unen para formar el holón H1 del nivel siguiente y cómo un cierto informón α producido por H2 forma junto con otro informón β el informón α del nivel siguiente.

Los informones pueden tener una estructura fractal, en el sentido de que pueden ser considerados formados por varios informones, dependiendo del grado de generalización requerida por el holón que lo usa: para un cierto holón, un registro de una cierta base de datos X será un informón, mientras que para otro holón toda la ontología, incluyendo la base de datos X, será un informón. La estructura del informón (cómo puede ser dividido) es, por lo tanto, dependiente del holón y no es algo intrínseco a él: por ejemplo, para un holón dado, la fecha “yymmdd” es solo un grupo de seis caracteres mientras que para otro es la concatenación de tres campos: “yy”||”mm”||”dd”; es decir, la estructura depende del marco de interpretación adoptado por cada holón, el cual a su vez sería un caso particular del marco usado por el holón “superior” en la jerarquía.

Un holón usa siempre al menos un informón (la descripción del estado del sistema en el que opera) y, a su vez, un informón debe ser usado por al menos un holón para tener

et al., 2018). Los autores definen una operación universal $\diamond$ que puede ser aplicada entre holones e informones, en cualquiera de las combinaciones, a saber:

- $H \diamond H \rightarrow H'$: es la integración de las funciones que realizan dos holones, para formar uno de mayor nivel $H \diamond H \rightarrow I'$: es considerar un holón como informón, tal como cuando se pasa a analizar el código fuente de un holón que es un agente inteligente.
- $H \diamond I \rightarrow I'$: Generación de nuevos informones por parte de un holón, como cuando se realiza fusión de información de varias fuentes (cada una proveyendo informones) para obtener un informón nuevo.
- $H \diamond I \rightarrow H'$: Mediante un informón se genera un nuevo holón, como en el caso de la compilación de un código fuente de un agente semi-autónomo para generar un nuevo holón, o cuando en un sistema que se auto-organiza cierta información produce reorganización en él.

2.8.3. Entidades

Según Sulis una entidad es un generador de informones (Sulis *et al.*, 2003). Las entidades generan colecciones de informones que forman historias dinámicamente coherentes. Algunos de los informones generados pueden ser hechos explícitos a través de la observación e incluso medidos, son los informones públicos. En este trabajo solo se consideran este tipo de informones.

Las entidades imponen un primer nivel de significado a los informones generados: el significado es que los últimos fueron generados por primeros. Un marco semántico es un sistema conceptual, un principio organizacional (Sulis *et al.*, 2003) que analiza y agrupa fenómenos en diferentes entidades, modos de ser, modos de comportamiento, modos de actuar e interactuar y asigna significado a todo ello en una forma coherente y consistente. El marco semántico modela y guía las interacciones entre las entidades y esos fenómenos que permiten mantener el hilo del significado; es una construcción primitiva tal como la noción de conjunto y más que describir un proceso particular, describe una coherencia significativa de procesos.

Dado el enfoque holónico de este trabajo, las entidades serán holones y cada holón interpreta (asigna significado a los informones usando distintos marcos semánticos).

2.9. Fusión de información

2.9.1. Definición

Con el tiempo han aparecido varias definiciones de fusión de datos (o de información, IF, *Information Fusion*, por sus siglas en inglés), la mayoría derivadas de los campos de detección remota (*remote sensing*) y militar. Por ejemplo algunas de las definiciones que citan Boström *et al.* (Boström *et al.*, 2018) son:

- Llinas (1988) “La fusión puede ser definida como un proceso de integrar información de múltiples fuentes para producir los datos unificados más específicos y comprensivos sobre una entidad, actividad o evento. Esta definición tiene algunas palabras claves operativas: específica, comprensiva y entidad. Desde un punto de vista de la teoría de información, para que sea efectiva la fusión como una función de procesamiento de datos debe (al menos idealmente) aumentar la especificidad y comprensión del entendimiento que tenemos sobre una entidad del campo de batalla o sino no tendría sentido ejecutar la función” (Llinas, 1988).
- Richardson and Marsh (1988): “La fusión de datos es el proceso por el cual los datos provenientes de una multitud de fuentes son usados para dar una estimación óptima de un vector de estado específico perteneciente al sistema observado.” (Richardson & Marsh, 1988).
- McKendall and Mintz (1988) “...el problema de la fusión de sensores es el problema de combinar múltiples medidas provenientes de sensores en una única medida del objeto o atributo percibido (sensed), llamado el parámetro” (McKendall & Mintz, 1988).
- Waltz and Llinas (1990): “Este campo de tecnología ha sido llamado apropiadamente fusión de datos porque el objetivo de sus procesos es combinar elementos de datos en bruto provenientes de diferentes fuentes en un único conjunto de información significativa que es de mayor beneficio que la suma de las partes contribuyentes. Como una tecnología, la fusión de datos es actualmente la integración y aplicación de varias disciplinas tradicionales y nuevas áreas de la ingeniería para alcanzar la

fusión de datos” (Waltz & Llinas, 1990).

- Abidi and Gonzalez (1992): “La fusión de datos tiene que ver con la combinación sinérgica de información hecha disponible por varias fuentes de conocimiento tales como sensores, de forma de proveer un mejor entendimiento de una escena dada.” (Abidi & González, 1992).
- Evans (1998): “La combinación de datos de varias fuentes complementarias (usualmente geo-demográficas y de estilo de vida o de investigación de mercado y de estilo de vida) para ‘construir una imagen de la vida de alguien’” (Evans, 1998).
- Wald (1999): “La fusión de datos es un marco formal en el que son expresados los medios y herramientas para la alianza de datos que se originan en distintas fuentes.” (Wald, 1999).
- Steinberg, Bowman and White (1999): “La fusión de datos es el proceso de combinar datos para refinar las estimaciones y predicciones de estado.” (Steinberg *et al.*, 1999).
- Gonsalves, Cunningham, Ton y Okon (2000): “La meta principal de la fusión de datos es combinar datos de múltiples fuentes en información que es de mayor beneficio que el que podría haber sido derivado de cada una de las partes contribuyentes.” (Gonsalves *et al.*, 2000).
- Hannah, Ball and Starr (2000): “La fusión es definida materialmente como un proceso de mezclado, generalmente con la aplicación de calor para derretir juntos los constituyentes, pero en procesamiento de datos una forma más abstracta de unión o mezcla es usada. El ‘calor’ es aplicado con una serie de algoritmos que, dependiendo de la técnica usada, dan una relación más o menos abstracta entre los constituyentes y la salida terminada” (Hannah *et al.*, 2000).
- Llinas (2001): “La fusión de información es un proceso de información que se ocupa de la asociación, correlación y combinación de datos e información de múltiples sensores o fuentes para alcanzar estimaciones refinadas de los parámetros, características, eventos y comportamientos para entidades observadas en un campo observado de vista. Es implementado a veces como un proceso totalmente automático o como

un proceso de ayuda a los humanos para el análisis y/o soporte a la decisión”(Hall & Llinas, 2001).

- Challa, Gulrez, Chaczko and Paranesha (2005): La fusión de datos provenientes de múltiples sensores “es un componente central de los sistemas de percepción (sensing) interconectados que es usada tanto para – unir/combinar información complementaria producida por sensores para obtener una imagen más completa o para – reducir/gestionar la incertidumbre usando información de los sensores desde múltiples fuentes.” (Challa *et al.*, 2005).
- Jalobeanu and Gutiérrez (2006): “El problema de fusión de datos puede ser planteado como el cómputo de la función de distribución de probabilidad posterior de un único objeto desconocido dadas todas las observaciones” (A. Jalobeanu, 2006).
- Mastrogiovanni et al (2007): “La meta del proceso de fusión de datos es maximizar el contenido útil de la información adquirida por fuentes heterogéneas de forma de inferir situaciones relevantes y eventos relacionados con el entorno observado.” (Mastrogiovanni *et al.*, 2007).

Algunas de estas definiciones siguen siendo totalmente válidas en la actualidad, por ejemplo Mönks (Mönks, 2017) se adhiere a la de Steinberg, Bowman and White (Steinberg *et al.*, 1999) y dice que al realizar fusión de información «conocimiento nuevo o más preciso sobre cantidades físicas, eventos o situaciones es creado por la utilización de diferentes fuentes de información».

Por fusión de información (IF) se entiende aquí al proceso de combinar datos de distintas fuentes para obtener información que sea más significativa y útil que la que se puede obtener de cada fuente en forma individual, sin fusionar (Foo & Ng, 2013).

Un estudio más detallado de la IF sobre todo a nivel de sensores se puede ver en (Khaleghi *et al.*, 2015) y (Mönks, 2017).

2.9.2. Generalidades

Mönks Mönks (2017) distingue entre fusión directa e indirecta:

- Fusión directa: la combinación de fuentes múltiples del mismo tipo para aprovechar las ventajas estadísticas de un número de fuentes acrecentado, y de los efectos de redundancia que compensan el ruido o las fuentes defectuosas.
- Fusión indirecta: la combinación de múltiples fuentes de distintos tipos para incorporar información, la que está disponible en algunas fuentes pero no en otras en el proceso de fusión.

La fusión directa puede ser vista como un caso especial de la indirecta, donde la información está presente en todas las fuentes.

En las últimas décadas, las agencias de gobierno, universidades e industrias han realizado grandes esfuerzos para desarrollar varias implementaciones de sistemas de fusión de información o algunos componentes de ellos usando una diversidad de arquitecturas, sensores, plataformas, aplicaciones y lenguajes de programación. Ya que el campo de la fusión de información está creciendo a partir de la investigación en distintas áreas de aplicación, varias definiciones y términos asociados están incompletos. La mayor parte de la literatura de fusión de datos y los esfuerzos puestos en ella se enfocan en:

- 1) Técnicas de fusión de información y algoritmos asociados para integrar datos de colecciones homogéneas o no de sensores que pueden ser potencialmente inexactos redundantes o defectuosos y
- 2) Modelos estructurales y funcionales que implementen los sistemas de sistemas de fusión de datos.

Sin embargo, y más allá de la enorme cantidad de actividades de investigación realizadas, la fusión de información es todavía un campo de investigación relativamente nuevo donde la metodología básica esta todavía estructurándose, es más, donde hay todavía distintas definiciones de lo que es la IF como se vio.

El campo de fusión de la información está evolucionando rápidamente tal vez empujado por los recientes avances en *business analytics* para poder manejar los problemas de Big Data (Das, 2013). Se prevé que *business analytics* y la fusión de información

convergerán en un futuro. Algunas compilaciones de conceptos, modelos y herramientas (técnicas, métodos y algoritmos) para fusión de datos puede encontrarse en varios libros: (Bossé *et al.*, 2007; Liggins *et al.*, 2008; Hall & Jordan, 2010; Blasch *et al.*, 2012a; Mönks, 2017)) y sobre *business analytics* en (Das, 2013).

Las raíces tradicionales de la fusión de información tal como se la conoce actualmente están en la fusión de sensores; basta ver las definiciones anteriores de IF, donde las “fuentes de datos” son sensores y los “aspectos de interés en el entorno” son objetos en movimiento, cada uno representado típicamente por un conjunto de vectores de estado. Pero las demandas en la comunidad de fusión de datos están comenzando a exceder este pequeño intento en al menos dos aspectos (Lambert, 2009). Primero, expandir los “aspectos de interés del entorno” más allá de las consideraciones de rastreo de objetivos militares para incluir temas relacionados con: biografía, economía, sociedad, transporte y telecomunicaciones, geografía y política. Segundo, las “fuentes de datos” están excediendo los sistemas de sensores militares, para incluir: sistemas de comunicación, bases de datos, sitios web, medios públicos de comunicación, fuentes humanas, etc. Es precisamente en estas fuentes de datos en las que el campo de “analytics” ha emergido en las ciencias civiles. Tanto las “analytics” de naturaleza inductiva como la fusión de información, más deductiva (sintética), necesitan ser combinadas. Son complementarias y para soportar realmente la toma de decisiones en entornos dinámicos complejos ambas partes (análisis-síntesis) son esenciales (Paggi *et al.*, 2014).

El proceso de IF típicamente tiene varios niveles (Raol, 2015):

- *de nivel de datos*, que combina los datos de distintas fuentes (esencialmente homogéneas) para producir un nuevo conjunto de datos más útil e informativo que el inicial.
- *de nivel intermedio*, que combina los datos obtenidos antes (en el nivel previo, de datos) para obtener un “mapa de características” que permite distinguir objetos y relaciones de mayor nivel.
- *de alto nivel*, que combina las decisiones de varios expertos individuales usando para ello votación, lógica *fuzzy* y métodos estadísticos.

A su vez, por fusión de información holónica se entiende, dicho simplemente, a la realizada mediante una holarquía (sección 2.9.4).

Desde un punto de vista teórico y considerando los conceptos de la Sección 2.8 se puede decir que cuando un holón desea la mejora de la calidad de la información que tiene, puede hacer fusión de información de su propia información (informones) con informones adicionales (que pueden ser por ejemplo las respuestas a consultas hechas a otros holones).

2.9.3. El estándar JDL

Se han propuesto muchos estándares (o conceptualizaciones) para la realización de la fusión de datos, comenzando en los años 80 con el Intelligence Cycle (Gad & Farooq, 2002), el Boyd Control Loop (Das, 2008) y el Joint Directors of Laboratories (JDL) (Blasch *et al.*, 2012c) hasta los propuestos más recientemente: Object-centered information model (Kokar *et al.*, 2000), Extended OODA P (Hall & McMullen, 2004), el unified *data fusion* (λ JDL) (Laxhammar *et al.*, 2009) o el Dynamic OODA loop (Brehmer, 2005). Algunos métodos se han desarrollado específicamente para una aplicación particular, por ejemplo: para evaluar las amenazas del malware (Dai *et al.*, 2016; Gao *et al.*, 2017; Lan *et al.*, 2016; Mnatsakanyan *et al.*, 2012; Zhou *et al.*, 2018). De los modelos mencionados, el JDL es el prevalente, el más usado al día de hoy (Paggi *et al.*, 2014). El modelo JDL fue propuesto para categorizar las funciones relacionadas con fusión de datos. El modelo divide el proceso de fusión en etapas y hace corresponder una capa (nivel) del modelo a cada una.

Los niveles en el Modelo JDL fueron originalmente el resultado de un esquema de partición basado en los efectos combinados e interdependientes de: a) niveles de abstracción cambiantes (datos-información-conocimiento); y b) niveles variables de la complejidad del espacio de problemas. Los niveles están definidos como:

- **Nivel 0:** Estimación de los estados de las entidades que no llegan a ser un objeto (por ejemplo señales o características).
- **Nivel 1:** Estimación de los estados de objetos físicos discretos (por ej. vehículos, edificios).

- **Nivel 2:** Estimación de las relaciones entre entidades (por ej. agregados, intentos, oportunidades).
- **Nivel 3:** Estimación de impactos (por ej. consecuencias de actividades amenazantes en los activos y metas propios).
- **Nivel 4:** Refinamiento de procesos (fue inicialmente reconocido dentro de la versión de 1999 del JDL (Steinberg *et al.*, 1999): adquisición y procesamiento de datos para soportar los objetivos de detección (por ejemplo control de procesos de fusión y diseminación de los sistemas de información).

Durante la última década, se hicieron revisiones subsiguientes al modelo y se propusieron niveles adicionales (Blasch & Plano, 2005) para incorporar nuevo entendimiento de los temas de fusión de información para culminar en el modelo propuesto por el Data Fusion Information Group (DFIG) (Blasch *et al.*, 2012b). En el modelo DFIG se proponen dos niveles adicionales al modelo JDL original: **Nivel 5** – refinamiento del usuario y **Nivel 6** – gestión de la misión (Blasch *et al.*, 2012b). El Nivel 4 de fusión es conocido como Refinamiento del Proceso en la primer versión del modelo JDL. El proceso implica gestión de recursos para mejorar los resultados obtenidos en los niveles más bajos de fusión de datos. En la segunda revisión del modelo JDL (Blasch *et al.*, 2013), se extienden los niveles con el objetivo de incluir niveles duales de gestión de recursos. Además, otros autores han propuesto una redefinición del nivel 4 de manera que las funciones de nivel 4 estiman las medidas de rendimiento del sistema y las medidas de efectividad. La evaluación del rendimiento en la práctica es bastante problemática. Ng y Foo (Foo & Ng, 2013) muestran una tabla con algunos enfoques para la evaluación del rendimiento para sistema de fusión de datos distintos dominios de aplicación. Como se describirá más adelante en esta Tesis, la introducción de este estándar resulta de importancia, ya que el modelo propuesto utiliza el nivel 4 del estándar JDL, reflejando este documento evidencia empírica de dicho uso. Luego de varios años de investigación intensiva, la fusión de bajo nivel se ha vuelto un campo relativamente maduro por lo que el foco de interés se ha movido hacia la fusión de mayores niveles. Ng y Foo (Foo & Ng, 2013) demuestran esto a partir de la cantidad de publicaciones hechas en conferencias especializadas (ediciones de la conferencia FUSION). Una forma sencilla de comprobar esto es realizar una

búsqueda en Google del término “high-level information fusion”, que dió un resultado de 167.000.000 coincidencias, mientras que la del término “low-level information fusion” arrojó sólo 28000.

2.9.4. Fusión de información holónica

Dada la estructura del modelo JDL, surge el planteamiento de cómo realizar la fusión de información según el modelo de la mejor forma, ya sea en cuanto a la precisión, velocidad, eficiencia del uso de los recursos, etc. Teniendo en cuenta la estructura en capas del modelo, y tratando de mantener las condiciones de semi-autonomía de los agentes que realizan la fusión a cada nivel, la fusión realizada a través de una holarquía aparece como la más natural y se han propuesto dos trabajos al respecto (Paggi *et al.*, 2014; Solaiman *et al.*, 2015). Las diferencias básicas entre estos enfoques se muestran en la tabla 3:

<i>(Paggi et al., 2014)</i>	<i>(Solaiman et al., 2015)</i>
Las funciones del nivel 4 están embebidas en cada nivel, y son cumplidas por holones propios de ese nivel. Esos holones (uno por cada nivel del modelo) se eligen por negociación, la holarquía emerge incluso en sus aspectos de gestión.	Establece una jerarquía de células de fusión de información pero no especifica cómo se gestionaría el nivel 4. Más bien son jerarquías de células de fusion. Las funciones del nivel 4 del modelo son llevadas a cabo, de todos modos, por un holón que se encarga de la gestión, posiblemente de todos los niveles. La gestión es impuesta jerárquicamente.

Tabla 3: Diferencias entre los dos principales enfoques de fusión holónica

2.10. Auto-organización

En este trabajo de Tesis se entenderá por auto-organización al proceso (y a la característica que debe tener ese proceso) por el cual alguna forma de orden emerge sin

requerir ningún control externo y a partir de interacciones locales entre partes de un sistema inicialmente desordenado.

En el caso de los sistemas dinámicos, la teoría del caos enfoca la auto-organización como conjuntos dentro del espacio de estados donde es posible la predicción, al menos para un horizonte de tiempo dado. La auto-organización suele estudiarse asociada a sistemas donde existen interacciones sociales, tales como la sociología y los sistemas multi-agente.

En el último caso, gran cantidad de trabajo se ha hecho para estudiar cómo se puede lograr que el sistema presente capacidades de auto-organización (Di Marzo Serugendo *et al.*, 2005). En el caso de redes, las que se auto-organizan típicamente son las *small-world networks* (Watts & Strogatz, 1998) y las *scale-free networks* (Barabási, 1999). En ambos casos, las estructuras emergen a partir de interacciones de abajo-arriba (*bottom-up*) a diferencia de interacciones arriba-abajo (*top-down*) propias de las redes jerárquicas (Clauset *et al.*, 2007).

La capacidad de auto-organización es una característica deseable de un sistema. Como se verá en el Cap. 7, el sistema que implementa el modelo propuesto en esta Tesis, posee propiedades de auto-organización. Por ejemplo (Karageorgos, 2011) dice que “para sistemas que son complejos y operan en un entorno dinámico, el uso de auto-organización ofrece ventajas significativas tales como escalabilidad aumentada y robustez y reducción en el costo de comunicación y de unidad de procesamiento”.

Cameron *et al.* (Cameron *et al.*, 2015) sostienen que “los resultados obtenidos demuestran la efectividad que una arquitectura dinámica con control-auto organizable puede tener en las grillas de distribución modernas y futuras que usen un gran número de agentes distribuidos”.

Haken y Portugali aclaran: “el proceso de auto-organización tiene lugar en sistemas complejos abiertos que adquieren estructuras funcionales o espacio-temporales sin recibir órdenes específicas del exterior”. Se dice que son “complejos” en el sentido de que están compuestos de varios componentes que interactúan, partes, elementos, etc. y “abiertos” en el sentido de que intercambian con su entorno materia, energía e información. La información que intercambian puede ser pragmática, como el impacto del mensaje en los que lo reciben, semántica como el significado que lleva el mensaje o de Shannon, como una medida de la capacidad de un canal a través del cual pasa un mensaje. El

modelo desarrollado en esta Tesis se enfoca en la información en el sentido semántico y pragmático. Claramente el concepto de complejidad es subjetivo y puede ser solo definido con referencia a un observador. Haken y Portugali, sostienen que para que haya auto-organización debe haber intercambio de energía e información y/o materia con el entorno (es decir, el sistema debe ser abierto) (Haken & Portugali, 2017).

2.10.1. Características básicas de la auto-organización en el sistema

Es difícil definir criterios rígidos para la existencia de auto-organización y poder aplicarlos rigurosamente. En (Meer & Koppen, 2005a) clasifican las características que debe poseer un sistema para ser llamado auto-organizado en dos tipos: las básicas y las propias de la autonomía. No todas estas características deben darse a la vez en un sistema para que exista auto-organización; por ejemplo, Napster (Napster, n.d.) no tiene precauciones para evitar la sobrecarga de servidores ni contra ataques de DoS (*Denial of Service*, en inglés, o negación de servicio, en español) o la inyección de datos corruptos. De la misma forma, ni la existencia de aleatoriedad ni la aparición de una criticidad auto-organizada existe en Gnutella (Frankel & Pepper, 2000) y sin embargo son sistemas reales totalmente válidos⁵. A continuación se detallan esas características relacionadas con la auto-organización (tomado de (Meer & Koppen, 2005b)).

2.10.1.1. Características básicas

Las características de este tipo son autodeterminación de límites, reproducción, mutabilidad, organización, existencia de métricas y adaptabilidad, como se detalla a continuación.

- **Autodeterminación de límites:** las fronteras del sistema son determinadas por él mismo, la decisión sobre la afiliación de sus nuevos componentes debe ser hecha

⁵"La criticidad autoorganizada es un concepto introducido por Bak *et al.* (Bak *et al.*, 1987) para explicar el patrón de comportamiento de diversos sistemas o fenómenos complejos existentes en la naturaleza. Estos sistemas complejos evolucionan hacia estados "críticos", donde pequeñas alteraciones que se producen sin la intervención de ningún agente externo, pueden desencadenar una serie de eventos de cualquier tamaño conocidos como avalanchas. Estas avalanchas se dan a todas las escalas a las que se observa el sistema, lo que determina una de las principales características de los sistemas auto-organizados críticamente: la invariancia de escala." (García-Marín *et al.*, 2008).

por el propio sistema.

- **Reproducción:** el sistema puede y debe reproducir su estructura (mediante la adición, remoción o cambio de un *peer*, sus datos o relaciones o conexiones con otros *peers*).
- **Mutabilidad:** debe ser capaz de cambiar su estructura, por ejemplo cambiando las conexiones o formando clusters.
- **Organización:** el sistema se organiza ya sea como jerarquía, heterarquía o ambas.
- **Métricas:** es capaz de detectar perturbaciones disparadas desde el entorno, por ejemplo: falla de un *peer* o conexión -en este caso, cuando no responde el otro *peer* puede ser considerado una falla -, sobrecarga o manipulación de datos (envío de datos falsos).
- **Adaptabilidad:** es capaz de reaccionar a las perturbaciones apropiadamente (re-estructurando *peers*, incorporando redundancia, etc.).

Características para la autonomía

Cuando un sistema posee las características siguientes puede adaptarse autónomamente y desarrollar una «vida propia»(Meer & Koppen, 2005b):

- **Feedback:** incluye los mensajes que los *peers* se envían entre sí. Un sistema P2P suele estar expuesto a *feedback* positivo y negativo por lo que la estructura del sistema cambia en forma balanceada.
- **Reducción de la complejidad:** desarrolla estructuras y oculta detalles del entorno para reducir la complejidad general (por ejemplo a través de la formación de *clusters* o la creación de otras entidades como los *active virtual peers* o de holones).
- **Aleatoriedad:** usa la aleatoriedad como un prerrequisito para la creatividad. Esto permite la creación de una nueva estructura con poco esfuerzo.
- **Auto-organizado críticamente:** un sistema con auto-organización se conduce a un estado de criticidad (*criticality*). Demasiado orden así como mucho desorden

deben ser evitados por procedimientos adecuados. Esto resulta en una mayor flexibilidad ya que el sistema puede manejar distintos tipos de perturbaciones. Demasiado orden sería una estructura jerárquica, demasiado desorden una P2P pura.

- **Emergencia:** muestra propiedades que no son propias de ningún *peer* o que pueden haber sido desconocidas en tiempo de diseño.

2.10.2. Características adicionales de redes P2P auto-organizadas

Meer y Koopen citan como características de las redes P2P con auto-organización :

- La igualdad de los *peers* contrarresta el riesgo de los únicos puntos únicos de falla.
- Un sistema P2P debería ser idealmente capaz de reproducir su estructura, no solo sus datos. Esa reproducción debería ocurrir automáticamente y no disparada manualmente o provenir desde afuera.
- La auto-organización va más allá de las propiedades deseables de flexibilidad, adaptabilidad y robustez. Incluye el uso de componentes al azar que permiten al sistema crear nuevas estructuras viables. En todos los casos el control ejercido externamente es limitado al mínimo.

CAPÍTULO 3

PLANTEAMIENTO DEL PROBLEMA Y PREGUNTAS DE INVESTIGACIÓN

Sea Γ una red abierta P2P (que puede pensarse como un espacio inteligente generalizado) formada por agentes, artificiales o no, que responde los mensajes enviados por un agente externo (al que llamaremos Ω o también Omega) en la que se realiza IF para disminuir el impacto de la vaguedad e incertidumbre que hay en los mensajes. La red no tiene ninguna estructura ni tipología en particular ni se especifica ninguna técnica en especial de IF de forma que sus componentes pueden ser considerados heterogéneos.

Por otra parte los componentes de la red tienen recursos (medidos como cantidad de mensajes disponibles o tiempo usado para responder) limitados.

Los mensajes se suponen compuestos de un número de campos posiblemente vinculados por relaciones de composición y donde las calidades de cualquiera de esas composiciones varían dependiendo de cada agente. Un campo puede formar (es decir, ser usado para determinar) más de un único campo. Por lo tanto la estructura de datos no es necesariamente un árbol. Cada agente puede, en teoría, componer (predecir o deducir) cualquier campo compuesto y puede consultar a cualquier otro agente respecto del valor de un campo compuesto.

En estas condiciones se tienen las siguientes preguntas de investigación (*research questions*, RQ):

RQ1. ¿Mejora el rendimiento (medido en base a los éxitos obtenidos y mensajes consumidos) en una red P2P con recursos limitados (por ejemplo Γ) en la que los agentes adoptan el modelo inteligente de comunicaciones con fusión de información propuesto en esta Tesis?

RQ2. ¿Surgen estructuras holónicas en la organización de los agentes cuando éstos adoptan mecanismos inteligentes propios del modelo propuesto en esta Tesis?

CAPÍTULO 4

TRABAJO RELACIONADO

En este capítulo se hace una revisión de los trabajos que se han enfocado en alguna de las facetas del problema planteado.

4.1. Uso de la fusión de información para reducir la incertidumbre y vaguedad

La fusión de información ha sido usada para reducir la indeterminación y otras “imperfecciones” en los datos y para mejorar los resultados que se obtienen a partir de ellos; es decir, para reducir el impacto negativo de dichas imperfecciones en la toma de decisiones (Nakamura *et al.*, 2014; Raol, 2015; Frank, 2008; Smith, 2011). Reducir el impacto negativo implica que el receptor debería tomar las mejores decisiones considerando el contenido semántico o pragmático asignado al mensaje recibido. Al respecto, Foo et al. afirman que algunas de las ventajas de hacer la IF incluyen: mejora en la exactitud de los datos así como reducción de la incertidumbre y ambigüedad en ellos, mejora en el reconocimiento de la situación (*situation awareness*, SAW) y mejora en las inferencias que conducen a una mejor toma de decisiones (Foo & Ng, 2013).

Una ventaja del enfoque de IF es que a cada agente se le permite razonar y manejar la vaguedad e incertidumbre en una amplia variedad de formas. Por lo tanto, es especialmente interesante para sistemas abiertos, como ocurre en esta Tesis en los que no

se puede garantizar la homogeneidad de sus componentes. En dichos sistemas abiertos, con el aumento sostenido en la conectividad entre sistemas y el advenimiento de IoT, la fusión de más y más información a partir de fuentes de datos heterogéneas ha ido ganando terreno en los últimos años. El enfoque de IF propuesto en esta Tesis es, de algún modo, bioinspirado, ya que se basa en el método “natural” que una persona utilizaría al manejar vaguedad e incertidumbre, ya que las personas consultan distintas fuentes y luego se procesa la información obtenida para alcanzar resultados con mayor certeza o de menor vaguedad.

Para hacer la fusión de los datos provenientes de los distintos agentes consultados se pueden usar distintos enfoques. Los métodos bayesianos son un formalismo que se puede aplicar cuando se usan probabilidades para representar la incertidumbre (Castanedo, 2013; Groen *et al.*, 2017). La teoría de Dempster-Shaffer (Fine, 1977) generaliza los métodos bayesianos y es capaz de representar el conocimiento incompleto, la actualización de creencias y de manejar explícitamente la incertidumbre. La teoría de Dezert-Smarandache (Smarandache & Dezert, 2004) es una mejora a la de Dempster-Shaffer, ya que es capaz de combinar formalmente todo tipo de fuentes de datos independientes aunque se enfoca primordialmente en la fusión de datos, cualitativos o cuantitativos, inciertos e imprecisos y altamente conflictivos. Sobre esta teoría hay una gran cantidad de investigación hecha; véase, por ejemplo, (Florentin Smarandache, 2015). Los métodos semánticos usan, no solo los datos, sino, además, información semántica suministrada por distintas fuentes (Castanedo, 2013). La fusión de datos con vaguedad puede aplicarse a la agregación de juicios expertos vagos o la agregación de opiniones usando conjuntos *fuzzy* continuos (Bai *et al.*, 2016).

La fusión de información en sistemas P2P orientada a la mejora de la calidad de datos se discute en (Zhao, 2009). Dicho artículo presenta un modelo de integración de datos orientado a la calidad de los mismos. Sus autores proponen una estructura multicapa para sistemas P2P pensada para registrar fuentes de datos y gestionar los metadatos de la información de forma global. A partir de eso se construye un modelo de evaluación de la calidad de datos basado el esquema de fuente de datos y el carácter estadístico de los mismos. Separan el sistema P2P en dos niveles: el nivel de gestión del esquema y el nivel de fuentes de datos. Cada nodo de gestión mantiene un grupo de nodos de fuentes

de datos almacenando sus esquemas y metadatos. Los nodos de gestión se comunican entre sí para compartir la información que manejan, por lo que la estructura general es jerárquica. Para elegir la fuente de datos se usan reglas heurísticas, como por ejemplo: elegir la fuente con más registros o la que tenga más esquemas de integridad.

Según el estudio de la literatura, también se pueden usar otras técnicas *ad hoc*, por ejemplo basadas en promedios aritméticos de los datos fuentes (si son cuantitativos), tal como se hace en (V K & Tharini, 2017) o la moda de ellos; la fusión a realizar depende de la aplicación. En esta Tesis se estudia el efecto que tiene en un sistema el elegir dos formas distintas la información a ser fusionada. Es por ello que los algoritmos del Capítulo 5 se refieren solo a “realizar la fusión”, es decir, la fusión es tomada como una función genérica.

La incertidumbre puede presentarse a diferentes niveles en un sistema: en los propios datos, en los campos/fuentes a consultar, etc. Es importante aclarar que en la presente Tesis se maneja únicamente la reducción de la incertidumbre intrínseca en los propios datos, haciendo fusión de información para ello, y se supone que no hay incertidumbre acerca de qué campos consultar o utilizar para obtener el valor de otro campo; el caso de incertidumbre en la descomposición o equivalentemente en qué fuentes/campos fusionar se propone como trabajo futuro. Los problemas asociados a esa descomposición, para disminuir la incertidumbre son discutidos por Kessel (Kessel, 2006). El trabajo de Kessel es un estudio sobre la dinámica de la incertidumbre: el beneficio de la reducción de la incertidumbre haciendo síntesis por fusión versus el aumento de la incertidumbre debido a las asociaciones equivocadas (es decir, que se fusionan fuentes que no deberían fusionarse). En esta Tesis el problema de la asociación se presenta cuando cada agente se plantea cómo se descompone un campo o qué campos se usan indirectamente para hacer la fusión. Kessel admite que la fusión puede ser hecha por una persona, un sistema experto, un sistema combinado hombre-máquina, etc. Si la fusión se refiriera sólo a la síntesis, la fusión siempre mejoraría la situación; sin embargo, no puede haber fusión sin asociación (selección de con qué hago la fusión). Kessel concluye que los problemas de asociación empeoran al crecer el número de oportunidades de confusión: si hay L campos que pueden ser confundidos, la probabilidad de que se haga una mala asociación es $P \leq (1 - 1/L)$ por lo que el rendimiento (confiabilidad de las inferencias hechas bajo incertidumbre) se

va a degradar concordantemente.

4.2. Dimensiones de la IQ consideradas

En cuanto a las dimensiones de IQ consideradas en artículos que describen problemas relacionados con el planteado en esta Tesis, en (Ngai & Gunningberg, 2014) se habla de cinco dimensiones para la calidad: precisión, actualidad, completitud, relevancia y autenticación (en inglés, *precision*, *timeliness*, *completeness*, *relevance* y *authentication* respectivamente). Precisión se refiere a la exactitud de la medida; actualidad mide el tiempo que pasa entre que se detecta y se reporta un evento; completitud mide la cantidad de datos obtenidos que son relevantes con respecto a la realidad; relevancia depende de los intereses del usuario, el contexto y las aplicaciones; y autenticación se asocia con la identidad de los usuarios, la autenticidad de los datos y la integridad de la comunicación. Su enfoque de mejora sólo se aplica a tres de ellas: *timeliness*, *completeness* y *relevance*. La mejora de la calidad se logra en forma adaptativa solo para *timeliness* al cambiar la tasa de muestreo y la tasa de *reporting* adaptativamente al entorno. Para las otras dos dimensiones la optimización se realiza en forma estática: para *completeness* se basan en la optimización de la distribución geográfica de los sensores que no se mueven en base a patrones de movilidad de los que sí son móviles. En cuanto a la relevancia (*relevance*), se cuidó su optimización en el momento de diseñar el sistema.

En (Hariharan *et al.*, 2011) consideran la IQ de la fusión de datos como una función de las IQ de cada uno de los datos, y asocian esa IQ a suma de las varianzas del error de los datos para distintos casos (distintos objetos a ser captados por los sensores), lo que otros asocian a *accuracy*, aunque su trabajo no toma esa métrica como fundamental y dan a entender que podría ser otra.

En (Hermans *et al.*, 2009) se propone un sistema para aplicaciones WSNs que permite evaluar la calidad de datos de un sensor para luego hacer la fusión basándose en esa calidad estimada. En cuanto a las dimensiones de calidad, se consideran *accuracy*, *consistency*, *timeliness* y *relevance*. *Accuracy* es la diferencia numérica entre la muestra y el valor real de lo medido; *consistency* se relaciona con cómo una muestra cumple con un modelo definido por el usuario, el cual dependerá de la aplicación; *timeliness* se refiere a si el

dato está siendo recibido por el nodo a tiempo; y *completeness* refleja si un nodo ha tomado la suficiente cantidad de muestras (o si ha recibido una fracción suficiente de una transmisión) como para reconstruir lo medido. Para ellos, la calidad de los resultados de la fusión es la mínima de las calidades de los componentes.

Ozdemir y Xiao (Ozdemir & Xiao, 2013) presentan un esquema de IF tolerante a fallos que elimina los datos generados por sensores que funcionan mal y que generan *outliers* (atípicos). De esa forma se reduce la cantidad total de datos transmitidos en la red. Si bien no lo plantean explícitamente, su noción de calidad de un dato es que éste no sea un *outlier*, lo que estaría relacionado con las dimensiones consistencia (*consistency*) y relevancia (*relevance*).

Las arquitecturas que consideran el contexto del usuario y de su entorno han sido llamadas “conscientes del contexto” (*context-aware*) (Huebscher & McCann, 2005). Por ejemplo, pueden considerar la ubicación, luz ambiental, ruido ambiental, etc. del usuario como datos de contexto. En (Huebscher & McCann, 2005) describen una arquitectura para la entrega de información de contexto en aplicaciones de este tipo. La aplicación define una función de utilidad que calcula un valor asociado a la utilidad de cierta información de contexto en base a distintos atributos de calidad de esa información (*precision* y *sampling rate*) de manera que la arquitectura entrega la información que resulta en una mayor utilidad. Se utiliza una capa de servicio de contexto que toma la información de los sensores (que son los proveedores de contexto) y para cada solicitud la aplicación elige el proveedor de contexto que da más valor a la función de utilidad de la aplicación. Según sus autores, cuando hay cambios en los proveedores de contexto, la adaptación es suave. En el caso de la arquitectura planteada en esta Tesis, si se considera que cada agente está ejecutando una aplicación, se puede pensar que hay una función de utilidad del contexto por agente (ya que se usan distintos pesos para los campos, eso mediría la importancia de ese campo si se lo considera como contexto del campo consultado). Los autores del citado trabajo reconocen distintos tipos de contexto, cada uno con distintas dimensiones de calidad, y manejan un directorio de servicios donde se registra qué información provee cada proveedor y sus atributos de calidad. Obsérvese que en su arquitectura hay distintas dimensiones de calidad para cada proveedor (sus autores no especifican cuales).

Rogova et al. utilizan el contexto para ayudar a la toma de decisiones (Rogova *et al.*,

2013) y señalan que la IQ puede relacionarse con distintas combinaciones de atributos de calidad, dependiendo de la aplicación. Distinguen tres categorías de calidad de información, teniendo cada una de ellas distintas dimensiones: calidad de la fuente de información, calidad del contenido de la información y calidad de la presentación de la información. Las dimensiones del contenido de la información son: accesibilidad, disponibilidad, relevancia, actualidad e integridad (mencionadas en la bibliografía como *accessibility*, *availability*, *relevance*, *timeliness* e *integrity* respectivamente). La *accessibility* y *availability* se refieren a que los usuarios puedan acceder la información efectivamente; *accessibility* se asocia al costo de ese acceso mientras que *availability* es binario, si un agente puede o no acceder a ella en cierto momento. La actualidad se puede medir como la utilidad de la información al momento de estar disponible. En cuanto a *integrity* se refiere a la incertidumbre (como conocimiento parcial del verdadero valor de los datos, generalmente medida como una probabilidad) y a la imprecisión (que a su vez incluye la *accuracy* o sea su grado de aproximación y el grado de conflictividad e inconsistencia con otra información, que tenga y que incluye el caso de datos *fuzzy*). Las dimensiones de las fuentes de información son nivel de pericia (*expertise*), *reputation* y *objectivity* para el caso de fuentes subjetivas (personas, hipótesis o creencias) o relevancia, credibilidad y confiabilidad (*relevance*, *credibility* y *reliability* respectivamente) para las fuentes objetivas (sensores, modelos automatizados - o sea que son independientes del sesgo humano). La relevancia se refiere a la medida en que la información que se reporta es de un objeto o evento que se refiere al objeto en consideración. La confiabilidad está relacionada con la calidad de las creencias o plausibilidad asignada a su respuesta por un modelo o agente humano. Finalmente la calidad de la presentación de la información incluye actualidad (*timeliness*) y completitud (*completeness*, o sea que el sistema que genera esa información pueda representar todos los estados de la realidad), interpretabilidad (*interpretability*, es decir en qué medida los usuarios pueden entender la información presentada) e inteligibilidad (*understability*, en qué medida el usuario es capaz de seguir la lógica del proceso automático que produce esta información).

4.3. Formas y objetivos de la fusión de datos

4.3.1. Clasificación de la IF según la topología

Las topologías mencionadas en la literatura (Chao & Hsiao, 2014; Dietzel *et al.*, 2009; Fan *et al.*, 2007) son: a) jerárquica o en árbol, b) agrupada en bloques (*clusters*) y c) totalmente distribuida. Se pasa a dar ejemplos de trabajos en cada una de ellas.

4.3.1.1. Jerárquica o en árbol

En una red de sensores no todos los nodos tienen por qué ser un sensor: se necesita que al menos un nodo reciba los datos para su procesamiento. Esto es, existen dos tipos de nodos: "fuente" de datos (asociados a un sensor) y nodos(s) "sumidero(s)" (*sink*) de datos. Ngai y Gunningberg (Ngai & Gunningberg, 2014) justifican el uso de un nodo sumidero porque la conexión 3G o WiFi era más cara en términos de energía y de equipamiento que una *bluetooth* u otras conexiones de corto alcance. Al día de hoy la diferencia, al menos en el sentido energético, permanece. En la fusión basada en árboles, los nodos de la red se organizan jerárquicamente y se construye un árbol para enviar los datos desde los sensores al nodo sumidero, el que los fusiona con sus propios datos (Villas *et al.*, 2013). La fusión se realiza cuando dos o más paquetes de datos llegan al mismo nodo del árbol. Los datos fluyen de las hojas del árbol hacia el nodo raíz, que corresponde al nodo sumidero (Dietzel *et al.*, 2009). Se puede decir que en la fusión jerárquica la red tiene sensores y centros locales de fusión. Los sensores envían luego información a los centros de fusión locales y allí se hace la fusión; los centros locales mandan luego la decisión al nodo raíz.

En (Huebscher & McCann, 2005) describen una arquitectura para la entrega de información de contexto en aplicaciones que tienen en cuenta el contexto (*context aware*). Dado que su modelo supone varios niveles, cada uno requiriendo servicios del componente del siguiente nivel (sensores, provisión de contexto, adaptación de la provisión de contexto y aplicación) el modelo sería jerárquico.

Según (Snidaro *et al.*, 2015) la mejora de la calidad de la fusión se obtiene usando el contexto, entendiéndose por contexto al “conjunto estructurado de variables, restricciones externas a algún proceso cognitivo (natural o artificial) que influencia el comportamiento

de ese proceso en el/los agentes en consideración” (Kandefer & Shapiro, 2008). En ese sentido, cuando un agente, en el modelo presentado en esta Tesis, pregunta a otro por un campo elemental, podría pensarse que ese campo es de contexto para poder formar los campos de los que forma parte. De forma más general, se puede decir que el contexto al hacer IF puede ser generalizado para establecer expectativas sobre el estado de entidades, eventos, etc. y resolver ambigüedad (Snidaro *et al.*, 2015). Los autores de ese trabajo mencionan a *Adaptive Middleware* como ejemplo de *middleware* jerárquico que utiliza contexto (Huebscher & McCann, 2005) (donde si bien no se menciona una topología, sí se aclara que en cierto nivel se agregan los datos de bajo nivel).

La fusión de información en sistemas P2P orientada a la mejora de la calidad de datos se discute en (Zhao, 2009). El artículo presenta un modelo de IF orientado a la calidad de datos. Proponen una estructura multicapa para sistemas P2P para registrar fuentes de datos y gestionar los metadatos de la información de forma global. A partir de eso se construye un modelo de evaluación de la calidad de datos basado el esquema de fuente de datos y el carácter estadístico de los mismos. Separan el sistema P2P en dos niveles: el nivel de gestión del esquema y el nivel de fuentes de datos. Cada nodo de gestión mantiene un grupo de nodos de fuentes de datos almacenando sus esquemas y metadatos. Los nodos de gestión se comunican entre sí para compartir la información que manejan, por lo que la estructura general es jerárquica.

Li y Pang (Li & Pang, 2018) estudian el caso de fusión desde múltiples fuentes heterogéneas, dado que su aplicación, que consta básicamente de un nodo donde se realiza la fusión de información proveniente de distintas fuentes, puede considerarse como un caso especial de jerárquica.

Las mismas consideraciones valen para (Guo *et al.*, 2018): estudian el caso de fusión desde múltiples fuentes heterogéneas. Su método utiliza información de varias fuentes (voz, imagen y vídeo que el usuario tenga en el celular, plataformas sociales, aplicaciones, GPS (Global Positioning System) para analizar los requerimientos de los consumidores para sistemas de recomendación de e-commerce. Emplean la información de la ubicación del comprador y también las características del comportamiento online de la persona. El artículo asume que toda la información que se precisa está disponible para ser recolectada. Si bien es una aplicación de topología típicamente centralizada, se la incluye como árbol

por ser la topología de “estrella” (centralizada) un caso particular de árbol en el cual no hay nodos intermedios entre las hojas y la raíz.

4.3.1.2. En *clusters*

Los enfoques basados en agrupamientos (*clusters*) consideran una organización jerárquica de la red, como los basados en árboles, aunque en este caso se distinguen dos tipos de nodos: los comunes y los cabeza de *cluster*. Los cabeza de *cluster* son elegidos para realizar la IF localmente y enviar el resultado al nodo sumidero (Villas *et al.*, 2013).

El modo de *clustering*, en especial aquel en que los *clusters* se determinan dinámicamente, sería el que correspondería con el modelo presentado en esta Tesis. Por ejemplo, en (Hermans *et al.*, 2009) se propone un sistema para aplicaciones WSNs que permite evaluar la calidad de datos de un sensor para luego hacer la fusión basándose en esa calidad estimada. Para realizar la fusión de los sensores se utiliza la teoría de Dempster-Shafer (D-S) y se devuelve el valor estimado de lo medido y su calidad (tal como se hace en el modelo propuesto en esta Tesis), sin hacer consideraciones de si la calidad devuelta es mayor que la del cabeza y sin usar nodos sumidero. El valor más probable de la cantidad medida se infiere a partir de un número de muestras correlacionadas basadas en su calidad estimada. Las cabezas de nodo se determinan eligiéndolas en cada *cluster*, aunque el *cluster* en sí viene dado. Es la cabeza del *cluster* la que hace la fusión usando la teoría de D-S.

Una red basada en *clusters* en la que para la formación de los mismos se tiene en cuenta el recurso energía es descrita en (Azharuddin *et al.*, 2015). El trabajo pone énfasis en el caso de que fallen los nodos cabezas de *cluster* y supone que luego de instalados los sensores ya no se mueven físicamente (es decir que permanecen fijos, lo cual no es el caso explicado en esta Tesis). Su algoritmo de formación de *clusters* trabaja en base a ciclos: descarta los datos redundantes y no correlacionados a nivel de un nodo cabeza y entonces el nodo cabeza envía los datos a otro nodo cabeza. En el siguiente ciclo se repite el proceso para los nodos cabeza de otro *cluster*. El artículo supone que al menos dos cabezas de cluster posibles están dentro del rango de transmisión de cada nodo y que luego que se instalan los sensores son estáticos (lo cual en la presente tesis no se supone).

Los nodos no tienen información global de la red.

La red propuesta en (Amgoth & Jana, 2015) también utiliza *clusters*. Se basa en la elección inteligente de los nodos cabeza de *cluster* a partir de lo que queda de energía de las cabezas de *clustering* y las distancias entre los nodos del *cluster*. Se permite la auto-organización de los *clusters*. Un nodo se autoproclama cabeza en base a la energía que le queda (se supone que los cabeza consumen más porque deben transmitir hacia afuera del cluster). Cada nodo debe conocer la energía residual de los nodos cabeza para poder decidir a qué nodo se integra (es decir, un dato global).

Existen otras redes basadas en *clusters* que consideran las fallas (mal funcionamiento) de los nodos. Por ejemplo, en la *Direct Diffusion* (Intanagonwiwat *et al.*, 2000) se crean múltiples caminos disjuntos entre los nodos y la base (Ω) y, si uno de los caminos se interrumpiese, se elige otro camino para transportar los datos. En caso de las redes que usan el algoritmo MHRM (*Mínimum Hop Routing Model*) (Chiang *et al.*, 2007) cada nodo envía sus datos a Ω de forma que el número de comunicaciones entre nodos (hops) es minimizado. En el caso de las que usan el CMATO (*Cluster-Member-bAsed fault-Tolerant mechanism*) (Lai & Chen, 2007) cuando falla un cabeza los miembros del *cluster* lo detectan y re-eligen una nueva cabeza para formar un nuevo *cluster*. Por lo tanto el sistema formado es auto-organizado.

Buscando la eficiencia en cuanto a la cantidad de datos transmitidos, lo que se hace en (Kamarei *et al.*, 2015) es tratar de reducir la cantidad de datos que reenvía la cabeza del *cluster*. Los *clusters* se forman jerárquicamente, siendo el 0 el nivel de los sensores, el 1 el de los nodos cabezas, el 2 el de los clusters de nodos cabezas, etc. Si $L(i)$ es la tasa de llegada de paquetes de datos al nivel i , lo que se hace es calcular un número $q = e^{-1/L(i)}$ y con probabilidad $1 - q$ se reenvía ese paquete en el nodo cabeza; en otro caso, se lo deja para fusionar con otros datos. Sus autores sostienen que así se reduce hasta el 70 % la cantidad de datos redundantes.

Asimismo, en (V K & Tharini, 2017) se utiliza un esquema de nodos cabeza de *cluster* donde se hace la agregación de los datos (fusión) usando para un algoritmo que trabaja en fases, en las que se forman clusters temporales para detectar un evento y en el que la cabeza del *cluster* hace la fusión de todos los datos y elimina *outliers* y datos redundantes, de forma que los datos resultantes son luego transmitidos a un nodo *sink*.

En la DRINA (Villas *et al.*, 2013) se busca aumentar la precisión de los datos y proponen una arquitectura basada en *clusters*: para cada nuevo evento se hace el agrupamiento (*clustering*) de los nodos que lo detectaron y se elige la cabeza de *cluster*. El algoritmo busca maximizar los puntos de agregación y usar menos paquetes de control para construir el árbol de direccionamiento (*routing*). Proponen un algoritmo cuyo principal objetivo es construir un árbol de direccionamiento con los caminos más cortos para conectar a todos los nodos fuentes con el sumidero a la vez que se maximiza la agregación de datos.

En (Ngai & Gunningberg, 2014) analizan la IQ en redes de sensores móviles. Los elementos de la red que se proponen como caso de estudio no son homogéneos; existen *brokers*, sensores y *smartphones*, cada uno con distintas capacidades. Los *brokers* son dispositivos que soportan varios estándares de comunicación, como WiFi, Bluetooth, ZigBee y GPRS (*General Packet Radio Service*). Los *brokers*, que están interconectados mediante una red sin restricciones en cuanto al tipo de topología, permiten distribuir mensajes a múltiples suscriptores. Su rol es mantener información de los clientes asociados a él y almacenar las suscripciones de los clientes. Los sensores son responsables de coleccionar datos del entorno y compartir eventos (por ejemplo valores *outlier*) a suscriptores interesados usando el *broker*. Un *smartphone* puede actuar como un cliente o un proveedor, suscribiéndose de forma pertinente por medio de un *broker*. Cada *broker* es responsable de un área geográfica y lleva registro de todos los sensores de esa área, por lo que puede ser considerado cabeza de *cluster*.

En (Ozdemir & Xiao, 2013) se presenta un esquema que maneja *clusters* cuya cabeza realiza la IF y es elegida dinámicamente (aunque sus autores no aclaran cómo se forman los *clusters* o eligen las cabezas).

En el caso de MidFusion (Alex *et al.*, 2008) se hace fusión en *clusters*: dada una aplicación, el conjunto de sensores que proveerán los datos, es elegido en MidFusion con el objetivo de seleccionar el mejor conjunto de sensores que maximizan el potencial de alcanzar las metas de la aplicación, asegurando suficiencia y eficiencia. Las aplicaciones sólo necesitan especificar los requisitos de información a MidFusion en lugar de los sensores en sí que van a proveer esa información.

Otros ejemplos de minimización del recurso energía por medio de *clusters* son las

arquitecturas LEACH (Singh *et al.*, 2016), PAMAS (Singh & Raghavendra, 1998), S-MAC9(Ye *et al.*, 2002), ASCENT (Cerpa & Estrin, 2004) y otros (ver (Alex *et al.*, 2008)).

4.3.1.3. Totalmente distribuidas o sin estructura

En esta categoría se incluyen los modelos que no hacen ninguna suposición sobre la topología de la red.

Un caso especial son los sistemas *peer-to-peer* que se auto-organizan y que resisten a cambios en el entorno; por ejemplo, en (Healing *et al.*, 2008) se describe una arquitectura orientada a servicios (*service oriented architecture*, SOA) y totalmente distribuida. En ella se monitorea permanentemente la red de forma de que cuando falla un recurso que provee un servicio se sustituye dicho recurso con la alternativa más cercana, manteniendo la capacidad global (en el modelo presentado en esta Tesis también se le sustituye, pero no hay un servicio dedicado al monitoreo). Cada nodo puede proveer un número de servicios que publica usando una capa de mensajes de tipo publicar/suscribir (*publish/suscribe*); junto con el servicio, se da una indicación de su calidad de manera que se puede elegir el proveedor con mayor calidad (tal como se hace en esta Tesis al indicar la calidad de la información provista) y el número de veces que dio una respuesta con una IQ mayor que la de quienes lo consultaron. Sus autores definen contexto como el conjunto de atributos que son externos a los requerimientos de la tarea pero que si embargo pueden influenciar el rendimiento de los proveedores. Por lo tanto un selector de servicio construye un modelo de cuán adecuado es cada proveedor en cumplir cada tarea particular en cada contexto. Esta idea puede ser extrapolada al modelo presentado en esta Tesis de forma que se podría dar a los *peers* más información de contexto (por ejemplo los otros campos que el *peer* consulta podrían ser considerados como contexto, lo que sería una mejora a hacer al modelo actual). Los selectores (los propios *peers*) pueden compartir toda o parte de la información sobre los proveedores de servicios para acelerar el aprendizaje (es decir, que en su modelo se comparte la experiencia que se tuvo con un proveedor de servicio). En el modelo propuesto en esta Tesis, la información de las experiencias anteriores es individual del *peer* y no se comparte. Para elegir un proveedor de servicios se utiliza un modelo de selección. El modelo de selección consiste de registros que representan *clusters*

de experiencia para servicios usados para tareas particulares y contextos y son usados para simplificar el espacio problema (o sea hay agrupamientos de experiencias, mientras que en el modelo aquí propuesto las experiencias quedan desagregadas).

La optimización del uso del recurso energía se estudia en (Hariharan *et al.*, 2011) considerando la IQ de la fusión de datos como una función de las IQ de cada uno de los datos fuente. En su modelo, se supone que existen una cierta cantidad de objetivos que pueden ser rastreados por conjuntos de sensores, y lo que se busca es asignar los objetivos a grupos (a determinar) de forma que la IQ obtenida para cada objetivo sea máxima considerando las restricciones de energía de cada nodo.

En (Kim *et al.*, 2017) se propone una arquitectura P2P auto-organizada para el monitoreo de la salud de personas que permite el envío en tiempo real de múltiples datos biológicos sin usar un servidor central al conectar al paciente con su cuidador, pero no realiza IF.

Snidaro *et al.* (Snidaro *et al.*, 2015) dan como ejemplo de *middleware* que utiliza contexto y que es distribuido a Nexus (Jakob *et al.*, 2006). Nexus tiene todos sus nodos en un mismo nivel y se le agregan tecnologías que permiten ofrecer una capa de mensajería de tipo *Publish/Subscribe* (*Pub/Sub*).

El contexto es usado también para evitar los problemas del conocimiento localizado en cada agente en sistemas multi-agente (F. Buford *et al.*, 2010). Sus autores presentan una arquitectura para sistemas de gestión distribuida de aplicación militar que combina IF, sistemas multi-agente y redes *peer-to-peer* y que puede estar formado por componentes heterogéneos.

Yousefi *et al.* (Yousefi *et al.*, 2012) se enfocan en protocolos que logren que los paquetes de datos lleguen al mismo tiempo al nodo que hace la fusión y que se pueda decidir a quién enviar la información fusionada. Suponen que los nodos se distribuyen al azar y utilizan un nodo sumidero que recopila los datos, no teniendo limitaciones en sus recursos mientras que los nodos son sensores a batería. Los sensores son iguales en cuanto a capacidades y limitaciones. Los nodos transmiten sus datos de un nodo a otro, hasta llegar al sumidero (a esa forma de comunicarse se la llama *hop-by-hop*) y dos nodos pueden ser o no adyacentes. En el modelo presentado en esta Tesis se aplica la convergencia temporal, es decir, se espera a que lleguen todas las respuestas a una consulta, aunque se mantiene un tiempo

de *time-out*.

La reducción del volumen de datos haciendo fusión se hace en algunos casos mostrados en la literatura usando razonamientos basados en lógica *fuzzy* y reglas para decidir cuándo dos datos son lo suficientemente similares como para fusionarlos, tal como ocurre en (Dietzel *et al.*, 2009). En dicho trabajo se presenta un mecanismo de fusión de datos independiente de la estructura de la red y que se basa sólo en su correlación. En su trabajo, la información sobre el tráfico en una carretera, después de ser fusionada por un vehículo, es enviada a todos los otros vehículos cercanos y a algunos lejanos.

Fan, Liu y Sinha (Fan *et al.*, 2007) buscan diseñar protocolos y técnicas que lleven a una fusión eficiente sin un mantenimiento explícito de la estructura de la red resaltando la importancia de la sincronización en la llegada de los mensajes a la vez para poder agregarlos.

Bertossi y Bravo (Bertossi & Bravo, 2004) se enfocan en los datos perdidos y las consistencias cuando la información está distribuida entre los *peers* y se usa la red para responder consultas. Estos autores utilizan la calidad de la información para manejar inconsistencias: se piensa en la información de las fuentes y en los *peers* para decidir entre las resoluciones posibles de inconsistencias.

4.3.2. Clasificación de la IF según las funciones usadas

Hariharan *et al.* (Hariharan *et al.*, 2011) plantean una metodología para hallar una solución al problema de maximizar la calidad total de la información obtenida, sujeta a restricciones de energía, usando una red de sensores que rastrea múltiples objetivos y en la que se pueden usar distintas funciones de fusión sin variar los resultados (no las especifican).

En Hermans *et al.* (2009) se propone realizar la fusión de información de un gran número de sensores con la teoría de Dempster-Shafer (D-S) y se devuelve el valor estimado y su calidad (tal como se hace en el modelo propuesto en esta Tesis), sin hacer consideraciones de si la calidad devuelta es mayor que la del nodo cabeza y sin usar nodos *sink*. El valor más probable de la cantidad medida se infiere a partir de un número de muestras correlacionadas basadas en su calidad estimada.

Buscando la eficiencia en cuanto a la cantidad de datos transmitidos, lo que se hace en (Kamarei *et al.*, 2015) es tratar de reducir la cantidad de datos que reenvía la cabeza del *cluster*. Usan para eso un método de baja latencia para agregado de datos (*Low-Latency Data Aggregation*, LLDA), para redes de sensores inalámbricos. Ese método opera de forma distribuida sin necesidad de procesar los datos, por lo cual no hay demora en la disponibilidad de los mismos. En realidad, más que agregado, lo que se hace es una selección y envío probabilístico de los datos.

En (Ozdemir & Xiao, 2013) hay un nodo “agregador” que recibe los distintos códigos H (códigos de tipo *hash* generados a partir de secuencias de m valores de los datos) de manera que les permite deducir si dos datos son iguales por lo que solo se transmiten esos H y le solicita sólo a algunos agentes que envíen sus datos - los que no son *outliers* ni redundantes- pero no especifican qué función usa.

Asimismo, en (V K & Tharini, 2017) se hace la agregación de los datos (fusión), la que puede hacerse usando distintas funciones simples (SUM, MAX, MIN, etc.). El algoritmo detecta *outliers* basándose en las lecturas de m valores de un sensor y en su valor *hash* H . Se eliminan los datos con igual código H y se supone que son un *outlier* en base a dicho código y el de sus sensores vecinos.

Dos trabajos estudian el caso de fusión desde múltiples fuente heterogéneas: (Guo *et al.*, 2018) y (Li & Pang, 2018). En el primero se utiliza un método que usa información de varias fuentes para analizar los requerimientos de los consumidores para sistemas de recomendación de *e-commerce*. Para la fusión de las distintas fuentes usan una versión de la teoría de Dempster-Shaffer. Para asignarle el peso a cada fuente usan una función de base radial (*Radial Basis Function*, RBF). Para definir la credibilidad de cada fuente se puede realizar un análisis estadístico o contar con la experiencia de un experto (juicio experto). En el segundo, se utiliza una versión modificada de la teoría de Dempster-Shaffer para predecir el riesgo de navegación (la probabilidad de que ocurra un accidente de tráfico marítimo). Sostienen que la teoría estándar de Dempster-Shaffer no arroja resultados confiables cuando los datos son conflictivos. Por ello para mejorar el rendimiento de la decisión se usa un modelo basado en máquinas de soporte vectorial (*Support Vector Machines*, SVM). El modelo con SVM usa los datos producidos por la aplicación de D-S.

4.3.3. Clasificación de la IF según los objetivos

Los objetivos que se han encontrado para realizar la IF en redes P2P son en su mayoría la reducción del volumen de datos transmitidos y por ende el ahorro de energía, y solo en algún caso específicamente para mejorar la calidad de la información, como se detalla a continuación.

En (Ngai & Gunningberg, 2014) se realiza la fusión para reducir los datos transmitidos y evitar enviar los datos que no cambian (o cambian menos). Esto los lleva a resolver el problema del mochilero repetidamente ya que, dada la capacidad limitada de transmisión que tiene cada nodo, se desea maximizar $\sum V_i C_i$ sujeto a $\sum C_i < C$ y $C_i < TC_i \forall i$, siendo C_i la capacidad de transmisión de datos del nodo i , TC_i la capacidad máxima de transmisión que puede tener el nodo i y V_i la importancia (variación) de los datos que se enviaría.

En (Kamarei *et al.*, 2015) se busca reducir la cantidad de datos que reenvía la cabeza del *cluster*. Obsérvese que la fusión hecha no es para mejorar ninguna calidad, sino para disminuir los volúmenes de datos transmitidos. Se supone en su trabajo que la red manejada es tal que sólo unos pocos sensores detectan un evento que sucede aunque teóricamente todos están habilitados para hacerlo.

En (Ozdemir & Xiao, 2013) se solicita sólo a algunos agentes que envíen sus datos (los que no son *outliers* ni redundantes) y no se aceptan datos salvo de ellos. Por lo tanto la IF se realiza en ese caso para eliminar *outliers* y datos redundantes.

Asimismo en (V K & Tharini, 2017) la cabeza del cluster hace la agregación de los datos (fusión) y la eliminación de *outliers* y de datos redundantes.

Otro trabajo que realiza la fusión para, básicamente, eliminar la redundancia de datos es DRINA (Villas *et al.*, 2013). En este trabajo no obstante se busca además aumentar la precisión de los datos.

Huebscher y McCann (Huebscher & McCann, 2004), usan la IF para determinar cuál es el contexto correspondiente a los valores medidos por una serie de sensores (o sea, determinar cuál es de una serie de contextos), aunque no profundizan mucho más sobre ello.

En (Yousefi *et al.*, 2012) se realiza la fusión de datos para disminuir los datos trans-

mitidos y así reducir la energía consumida, al igual que en (Fan *et al.*, 2007).

La fusión de información en sistemas P2P orientada a la mejora de la calidad de datos se discute además en (Zhao, 2009), donde se presenta un modelo de integración de datos orientado a la calidad de datos. Proponen una estructura multicapa para sistemas P2P para registrar fuentes de datos y gestionar los metadatos de la información de forma global. A partir de eso se construye un modelo de evaluación de la calidad de datos basado el esquema de fuente de datos y el carácter estadístico de los mismos.

El trabajo (Guo *et al.*, 2018) estudia el caso de fusión desde múltiples fuentes heterogéneas, y se realiza la IF para ayudar a la toma de decisiones, ya que es un sistema de recomendación. Asimismo en (Li & Pang, 2018) se utiliza la IF para ayudar en la toma de decisiones y predecir el riesgo de navegación (la probabilidad de que ocurra un accidente de tráfico marítimo).

4.4. Redes con recursos restringidos

En este apartado se comentan los distintos tipos de redes con recursos limitados que se encontraron, en relación a dichos recursos.

En (Ngai & Gunningberg, 2014; Hariharan *et al.*, 2011; Yousefi *et al.*, 2012) lo que se busca como ahorro de recursos es un ahorro en la energía, lo mismo se busca en (Ngai *et al.*, 2009) por medio de la reducción del tráfico de mensajes usando un algoritmo distribuido y liviano (las decisiones son hechas localmente por los nodos y no requiere mensajes de *feedback*).

Un algoritmo de *clustering* (es decir, para redes basadas en *clusters*) que tiene en cuenta el recurso energía es descripto y discutido en (Azharuddin *et al.*, 2015); también se tiene en cuenta el recurso de energía de los nodos en redes con clusters en (Amgoth & Jana, 2015).

En el caso de redes implementando el protocolo de comunicaciones Direct Diffusion (Intanagonwiwat *et al.*, 2000) que permite la comunicación entre el nodo *sink* y los demás nodos en redes no estructuradas, se busca el ahorro de energía en la red tomando en cuenta las fallas que se producen en sus nodos. de datos a partir de consultas mientras que en el del protocolo MHRM (Chiang *et al.*, 2007) cada nodo envía sus datos a Ω de forma

que el número de hops es minimizado (y por lo tanto el gasto de energía, aunque podría ser el de mensajes enviados). Además, cuando alguno de los nodos falla o queda sin batería, el protocolo permite la adaptación al cambio y encuentra un camino alternativo. El mismo fin de ahorro de energía se ve en el CMATO (Lai & Chen, 2007) el que a su vez permite la auto-organización ya que si falla un nodo cabeza, los miembros del *cluster* re-eligen una nueva cabeza para formar un nuevo *cluster*. También se busca el ahorro de energía buscando la eficiencia en cuanto a la cantidad de datos transmitidos (evitando los duplicados) en (Kamarei *et al.*, 2015) y en (Ozdemir & Xiao, 2013), en el último se busca además reducir los *outliers*. Los problemas de consumo de energía, datos redundantes, generación de *outliers* y falla de nodos también son analizados en (V K & Tharini, 2017) y en el caso de redes con el protocolo DRINA (Villas *et al.*, 2013).

En (Hassanzadeh & Stoleru, 2014) se toman como recursos limitados la memoria y la capacidad de cálculos, aunque no se los cuantifica y sólo se los considera para justificar tener una arquitectura distribuida

4.5. Oportunidades de investigación encontradas

Una vez realizada la revisión bibliográfica se han encontrado una serie de carencias o debilidades en los enfoques analizados, como se pasa a explicar. Primeramente, las dimensiones de IQ consideradas son muy pocas ya que se suelen manejar una o dos dimensiones más, aparte de precisión (o *accuracy*) y *timeliness*; segundo, con respecto al manejo de recursos, los trabajos se enfocan en la reducción del gasto energético y cuando consideran reducción de mensajes es sólo para reducir consumo de energía, sin pesar que, en algunos casos, la cantidad de mensajes enviados puede tener, de por sí, importancia, como en el caso de cobro por mensaje en móviles pre-pagos o cuando el ancho de banda es limitado (Wang & Hovakimyan, 2010); no se encontraron trabajos que manejen el concepto “recurso” en forma genérica. Tercero, considerando la cada vez mayor interacción entre personas y máquinas, no existe un modelo operativo de la calidad de información que permita integrar en su funcionamiento personas con máquinas, y que considere la vaguedad (*fuzzyness*) - típica de las comunicaciones humanas. Esto ocurre incluso en modelos que consideran explícitamente a las personas como fuente

de información. (Rogova *et al.*, 2013). Los modelos existentes tampoco consideran las posibles distorsiones que se producen en la información por ser de segunda, tercera, ... mano.

Considerando estos puntos de carencia existentes en la literatura, queda clara la necesidad de cubrir los objetivos planteados en la introducción por medio del modelo propuesto (descrito en el Cap. 5), en el que la calidad de la información (la cual incluye varias dimensiones tales como *accuracy*, *fuzzyness*, confiabilidad, etc.) es mejorada usando IF, y que, además, es calculable aritméticamente, por lo que su cómputo puede ser mecanizado. El modelo planteado, así mismo, considera los recursos en forma genérica.

Las aportaciones del modelo propuesto pretenden, tal y como se comprobará en la parte de experimentación, dar respuesta de manera positiva a las preguntas de investigación (o *research questions*, por su nombre en inglés) planteadas, avanzando así el estado de la cuestión existente.

CAPÍTULO 5

DESCRIPCIÓN DE LA SOLUCIÓN

La finalidad de este capítulo es mostrar cómo se da cumplimiento a la aportación A1 tal como es descrita en la sección 1.3.

En esta Tesis, cuando se hace referencia a agentes colaborativos (o más brevemente, “agentes” o “*peers*”) no se está suponiendo que son todos del mismo tipo. Por ejemplo, en el caso de una red de *soft sensors*, cada uno de ellos va a ser calculado por un agente que puede ser mecánico o no. Los agentes a su vez colaboran y así se pueden generar *soft sensors* de segundo, tercer, etc. nivel. Una combinación algorítmica de *soft sensors* es un *soft sensor* también ya que en última instancia se está haciendo una combinación matemática (en dos o más pasos) de los valores arrojados por los sensores “*hard*”. El modelo planteado en esta Tesis será descrito considerando la IQ obtenida y cómo un sistema usando dicho modelo se comportaría.

Como se adelantó, el modelo asume que los agentes no tienen una organización específica pero sí que pueden comunicarse libremente entre sí, por lo que se los considera formando una red P2P plana, cuyos individuos poseen recursos limitados. Ejemplos típicos de esto son: a) el caso de una comunidad cuyos integrantes colaboran para hallar la respuesta más adecuada a una consulta que se les plantea desde afuera (tal como un sistema de recomendación de hoteles u atracciones turísticas) donde los usuarios (*peers* de la red) pueden dar en tiempo real su opinión (respuesta) a una consulta de alguien sobre la conveniencia de una opción usando su *smartphone* que tiene batería y tal vez

número de mensajes limitados; b) un escenario donde un grupo de personas está ayudando a manejar una catástrofe y se comunican usando dispositivos que tienen limitaciones (batería, duración de las transmisiones, etc.) y donde se pueden incorporar nuevas personas en cualquier momento; o c) un escenario de un grupo realizando maniobras militares en terreno hostil, donde la cantidad de mensajes está limitada por consideraciones de seguridad por ejemplo y donde ningún integrante puede considerarse imprescindible para la misión.

5.1. Aspectos generales de la propuesta

En este trabajo se propone un modelo de sistema multi-agente que responde a los mensajes enviados por un agente externo (llámese Ω) usando fusión de información para minimizar el impacto de la vaguedad e incertidumbre presentes en las comunicaciones del sistema. Este modelo está basado en una red P2P no estructurada cuya topología cambia con la disponibilidad de recursos. Se llamará a este tipo e-P2P (e de evolutiva o *evolvable*). Los mensajes se suponen compuestos de un número de campos posiblemente relacionados por relaciones de composición y donde las calidades de cualquiera de tales composiciones (fusión) varían dependiendo de cada agente. Un campo puede formar (ser usado para determinar) más de un único campo. Por lo tanto la estructura de datos no es necesariamente un árbol. Cada agente puede, en teoría, componer (predecir o deducir) cualquier campo compuesto (con mayor o menor calidad) y puede consultar a cualquier otro agente sobre el valor de un campo compuesto. Los agentes son entrenados en algún momento para poder hacer esto, usando, por ejemplo, técnicas basadas en redes *neuro-fuzzy*.

El sistema cambia, evoluciona, a un modo operativo más inteligente dependiendo del número de mensajes sin responder dentro de un límite de tiempo preestablecido y un límite específico en el número de mensajes disponibles por agente. Este cambio puede tener lugar gradualmente, agente por agente, o bien cuando una señal indicando recursos limitados es enviada a todos los agentes. La arquitectura inicial es un sistema P2P plano (descentralizado y sin estructura (Vu *et al.*, 2010a), donde ni la calidad ni el costo de la comunicación, con excepción, en algunas ocasiones, del número de mensajes, son tenidos

en cuenta. En el caso de que el cambio no sea gradual, un agente externo determina el instante en el cual los mensajes empiezan a ser limitados, que es el mismo para todos sus miembros. Este agente no es modelado aquí.

Aunque la idea de costo de los mensajes generalmente implica distancia geográfica, no se considera aquí si un agente está físicamente próximo a otro. El tiempo de respuesta a cada mensaje se supone aleatorio para cada agente. Al comienzo de la operación del sistema, todos los agentes están comunicados entre sí. En el caso de cambio gradual, a medida que pasa el tiempo, un proceso de selección natural ocurre: cuando un agente observa que hay demasiados TOs o cuando recibe una señal de mensajes limitados, simplemente cambia a modo inteligente (o económico) y comienza a dirigir sus mensajes exclusivamente a los agentes que han respondido recientemente y ofrecido mejor calidad de la información. Con el transcurso del tiempo, los agentes que agotan su capacidad de comunicación (representada por carga de la batería, el crédito de los mensajes telefónicos pre-pagos, la posibilidad física de hacer una llamada en un entorno hostil, etc.) van a ir dejando de responder. En base a ello, otros son seleccionados entre los agentes que no han sido consultados últimamente porque no proveen las respuestas de mejor calidad, les lleva demasiado tiempo o ingresaron al sistema luego de la consulta). Por lo tanto, el sistema siempre cambia a modo inteligente debido tanto al número acumulativo de TOs o el evento señalando la limitación en el número de mensajes. Sin embargo, como se menciona en el planteo del problema, nuevos agentes entran al sistema con el tiempo, aunque se supone que la tasa de entrada y salida no es tan alta como para ser un caso de *churn* (Stutzbach & Rejaie, 2006).

Dos enfoques mutuamente excluyentes pueden adoptarse al diseñar un sistema con IF descentralizado: agentes competitivos o cooperativos. En el caso cooperativo, los agentes compiten para dar medidas independientes de la misma propiedad o característica de un objeto. Esto puede ser útil para entornos de sistemas críticos (tales como los sistemas aeroespaciales) que deberían ser resistentes a las fallas. Por otra parte la información provista por los distintos agentes en un sistema cooperativo no es redundante. El modelo de sistema propuesto es tanto competitivo como cooperativo. En el lado competitivo, cada agente puede, en teoría, predecir todos los campos compuestos. Se busca ofrecer una respuesta de calidad superior a las de los demás. En el lado cooperativo, sin embargo,

esto es hecho de manera de maximizar la calidad de la respuesta final de todo el sistema.

En un sistema distribuido, la confiabilidad es uno de los aspectos de la calidad. En este trabajo, la confiabilidad está representada básicamente por el número de agentes consultados, el número de agentes que respondieron y sus respectivas calidades (es decir, a igual calidad de predicción, la predicción hecha al preguntarle a varios agentes sería más confiable que una predicción por la cual nadie fue consultado); sin embargo, el número de agentes consultados por uno y el de los que le respondieron debería ser alto, lo que implica enviar más mensajes. Esto puede ser contraproducente. Este dilema es resuelto añadiendo un exponente en la fórmula de calidad para especificar cuán importante es verificar el número de mensajes de respuesta en cada consulta.

La aleatoriedad es una característica importante de este modelo (como cuando un agente elige los campos sobre los que consultará a otros, cuando elige los agentes que consulta, cuando decide si responder o no si la respuesta de menor calidad que la del que lo consulta, etc.). De esta forma se ayuda a que los sistemas instanciados basados en este modelo se auto-organicen creativamente (Meer & Koppen, 2005b).

Si se consideran los agentes del sistema como entes situados en un cierto espacio físico, se comprende que todo el conjunto forma un espacio inteligente.

En ese espacio inteligente los componentes se relacionan de acuerdo a un cierto modelo inspirado en la naturaleza y que utiliza

- a) redes neuronales
- b) ideas relacionadas con la comunicación humana, en especial cómo las personas y organizaciones dan sentido a la información recibida, con las teorías de *sense-making* de Dervin y de Weick y
- c) técnicas inspiradas en el templado simulado.

El modelo busca mantener la robustez y adaptabilidad frente a cambios en su integración y del entorno, manejando información vaga e imprecisa, a la vez que hacer un uso racional de los recursos (comunicaciones, batería, tiempo, etc.) de los componentes del IS.

Se ignora aquí el fenómeno de *churn*.

Se utiliza una métrica de IQ se puede aplicar para tanto datos numéricos como simbólicos. Dicha IQ se computa como un escalar usando formulas *ad hoc* que incorporan otras dimensiones además de la vaguedad e incertidumbre y que son fácilmente calculables por cada *peer* y, evitándose así que sea una sobrecarga importante de cómputo para el *peer*.

Entre las dimensiones consideradas para el cálculo de la IQ (para un agente α consultando otros agentes) se encuentran:

- El número de agentes consultados por α , asociado con la credibilidad y confiabilidad del valor resultante.
- El número de agentes que contestaron la consulta, junto con el número de agentes consultados, es representativo de las dimensiones de credibilidad, eficiencia y confidencialidad (Rafique *et al.*, 2012; ISO, 2005).
- La vaguedad e incertidumbre con la que responde α para cada ítem de información. Esos dos valores pueden ser determinados por el agente consultado (mediante introspección), negociada con α o asignada directamente por α . Se ha implementado la introspección (cada agente sabe la calidad con que puede predecir cada campo). La principal fuente de incertidumbre de α será el error de las RNA que use. La vaguedad es representada como se propone en la Sección 2.2, y la incertidumbre, usando probabilidades (como el porcentaje de error del valor de respuesta).
- El número de reenvíos de la consulta o sea la importancia dada por el agente a que la información sea de primera, segunda, tercera ... mano, es decir, si fue suministrada directamente por el agente consultado o no. Este podría ser bajo por haber consultado a un número pequeño de agentes, por ejemplo. Es más, intuitivamente no es lo mismo consultar a cuatro agentes a la vez que consultar a A, que consulta a B, que consulta a C, que consulta a D. Algunos riesgos (por ejemplo, de seguridad) podrían ser mucho mayores si hay más re-envíos. Esto se corresponde con las dimensiones de confidencialidad y eficiencia (ISO, 2005; Rafique *et al.*, 2012).
- La importancia relativa de cada una de las partes constitutivas de un ítem de información para α . Por ejemplo, dado un mensaje compuesto de varios campos, cuán importante es cada campo al agente que consulta.

- La variación en la calidad de las respuestas dadas por los agentes consultados. Se busca minimizar la incertidumbre, pero una gran variación en la calidad da menos calidad o certeza que una pequeña variación en la calidad. Esto se asocia a la dimensión de credibilidad (ISO, 2005; Rafique *et al.*, 2012).

La toma de decisiones multi-criterio se evita usando un escalar como calidad.

Por último, el *sensemaking* se refleja aquí en varias formas. Primero, cada agente formando parte de la red (la cual es vista como una organización) trata de aclarar los mensajes que le llegan consultando a otros miembros de la organización, que resultarán ser sus favoritos para cada elemento de datos (campo) gracias a la reputación que vayan ganando (o sea, gracias a la calidad con que responden). Esas respuestas actúan como “pistas” para interpretar el mensaje recibido y esos favoritos constituyen el “entorno” de cada agente. Segundo, como se ve, sería un enfoque más bien organizacional como el propuesto por Weick. La realidad, en el sentido del significado dado por el sistema a un mensaje, emerge continuamente a medida que los elementos de la red interactúan. Tercero, dado que no se especifica cómo realizar la fusión de las respuestas obtenidas y del valor propio del que consulta se puede tomar que existe cierto nivel de invención en la respuesta.

5.2. Fórmula de calidad usada

Sea un agente α que recibe un mensaje m con cierta calidad de parte del agente Ω (la calidad depende del agente y del campo en cuestión, aunque también se calcula la calidad del campo a partir de la de sus componentes). Existe una probabilidad de que α consulte a otros agentes por sub-campos de m , m_t (pudiendo ser $m \equiv m_t$, es decir, que se puede consultar por todo el mensaje) los que a su vez pueden repetir el proceso. Luego de tener todas las respuestas (o de haberse producido los TOs correspondientes), α hace la fusión de información obtenida pertinente. α recibe el número de veces que se reenvió a consulta para m_t (llamado aquí el “nivel” del agente para esa consulta) así como el número de agentes consultados.

Siempre que se hable de sub-campo se está refiriendo a sub-campo en un paso de descomposición (o sea, si A es sub-campo de B y B de C, A no es un sub-campo de

C). La descomposición de campos en ciertos sub-campos es propia de cada dominio de conocimiento, de cada aplicación, y se considera única (común) para todos los agentes.

La calidad del campo m_t para el agente α : $Q(\alpha, m_t)$ se calcula como sigue. Sea Z el parámetro que indica qué tan importante es que una respuesta sea de primera mano o no. Si no tiene importancia, se toma $Z = 0$ y si la tiene, $Z > 0$. A mayor Z , más importancia. Sea $\bar{Q}$ la calidad de α al predecir sin consultar a nadie. El valor Q = la calidad de α al predecir (estimar) se calcula como una función de la vaguedad y la incertidumbre para m_t en α si α no consulta a nadie o como una función que considera las vaguedades e incertidumbres de α y de los agentes consultados que respondieron, la cantidad de agentes consultados en total en forma directa por α y Z .

Analíticamente, para calcular Q = la calidad de α al predecir (estimar) m_t se toma:

Si α no consulta a nadie o si no le respondió nadie es:

$$Q(\alpha, m_t) \equiv \bar{Q}(\alpha, m_t) = \left(\frac{1}{[1+|H(\alpha, m_t)|][1+U(\alpha, m_t)]} \right) \quad (Ec.7)$$

En otro caso:

$$Q(\alpha, m_t) \equiv [\max\{\bar{Q}(\alpha, m_t), q\}]^{(Z+1)/R} \quad (Ec.8)$$

donde

$$q = \frac{1}{[1+V(\alpha, m_t)]} \max_{\beta_i \in T} \left\{ \frac{Q(\beta_i, m_t)}{L(\beta_i, m_t)^Z} \right\} \beta_i \in T \quad (Ec.9)$$

$$R = |T|^2 / S \quad (Ec.10)$$

siendo

- $\bar{Q}$ la calidad de α al predecir (estimar) m_t
- L es el número de reenvíos que tuvo la consulta (o sea de cada vez que se generó otra consulta a partir de ella). Se define $L(\alpha, m_t)=1$ si α es consultado por Ω y $L(\beta_k, m_j)=L(\alpha, m_t)+1$ si α consulta a $\beta_k \quad \forall j, k, t$.
- T es el conjunto de los agentes β_i consultados por α , $|T|$ denota su cardinal
- S el número de agentes consultados por α para m_t , hayan o no respondido

- V es una medida de la variación de las calidades de las respuestas recibidas por α para m_t . Por ejemplo podría ser

$$V = \max_{i \in T} \{Q(\beta_i, m_t)\} - \min_{i \in T} \{Q(\beta_i, m_t)\} \quad (Ec.11)$$

- $U(\alpha, m_t) \in [0, 1]$ es una medida de la incertidumbre propia de α sobre m_t , por ejemplo, el porcentaje de error cometido al calcular α a m_t (por ejemplo, por deficiencias de un sensor).
- $H(\alpha, m_t)$ es una medida de la vaguedad del dato

Luego calcular la calidad del campo m según α a partir de las calidades de sus componentes según la fórmula

$$Q(a, g) = \sum_j W_j(g_j) Q(\alpha, g_j) \text{ si es } g \text{ formado por } \{g_1, g_2, \dots, g_j, \dots\} \quad (Ec.13)$$

siendo $W_j(g_j)$ el peso (importancia) que le da el agente a g_j para formar g .

Los parámetros de estas fórmulas se corresponden con las dimensiones consideradas en 4.2 de la siguiente manera:

- a) La importancia relativa de cada campo, expresada por los pesos W en la ec. 12.
- b) La importancia dada a que la información sea de primera, segunda... mano dada por el exponente Z en la ec. 3 y el uso del parámetro L que indica cuántas veces se reenvió la consulta.
- c) La variación de la calidad de las respuestas, expresada por V en la ec. 9.
- d) El número de agentes consultados, asociado a S en la ec. 10.
- e) El número de agentes que contestaron, asociado al conjunto T en las ecuaciones 9 y 10.
- f) La vaguedad e incertidumbre en la respuesta de un agente a α esta dada implícitamente por Q en la ec. 10.
- g) La vaguedad e incertidumbre propias de α reflejada en H y U en la respuesta de α en la ec. 7.

h) La cantidad de reenvíos del query, expresados por L en las ec. 7 y 10.

La calidad de un campo es calculada por α como una combinación lineal de las calidades de sus campos componentes.

Los distintos niveles de IF (por ejemplo datos, medio y alto) son representados en el modelo por las relaciones de composición. Por ejemplo, la fusión de alto nivel corresponde al valor de respuesta al mensaje completo m y la fusión a alto nivel al valor predicho para un campo no compuesto.

5.3. Dinámica de un sistema con este modelo

En términos generales se puede decir que el espacio inteligente actúa como si fuese un sistema P2P que opera recibiendo mensajes de un agente externo (Ω) que interactúa con dicho espacio. El agente externo pide al sistema estimar ciertos campos del mensaje (o incluso el mensaje completo), elegidos al azar (en realidad Omega lo elige con algún criterio, que por ser desconocido para los demás agentes, es como si fuera al azar) por el agente consultante. Típicamente el o los otros campos a estimar serán usados para tomar una decisión. Se espera que el sistema retorne esta estimación o estimaciones dentro de un tiempo dado. Si este tiempo límite no se cumple, se dice que se produjo (o hay) un *time-out*. Es más, generalmente ni el sistema ni los agentes responden si su calidad es menor que la del sistema o agente que consulta (aunque esto puede no suceder con una probabilidad muy baja). Este comportamiento está inspirado en el recocido simulado.

Al comienzo todos los agentes reciben las consultas. Con el paso del tiempo, esto es, una vez que se ha determinado qué agentes proveen a Ω con información de mayor calidad para los campos especificados (o todo el mensaje), solo unos pocos (a lo sumo N^* – un parámetro del sistema – para cada campo o todo el mensaje) reciben las consultas.

A medida que se agotan los recursos de los agentes, éstos se vuelven incapaces de responder. Debido a ello nuevos agentes deben ser seleccionados para responder una consulta. Recursivamente, cada agente en el sistema puede consultar o ser consultado por la estimación de un campo.

Las Figuras 3-6 diagraman el comportamiento del sistema. En primer lugar ningún agente tiene preferencias por ningún agente por un campo. Todos los agentes reciben

una consulta (Figura 3), ya que Ω no tiene aún preferencias de agentes. Esta consulta es procesada y puede llevar a realizar otras consultas entre los agentes del sistema. A medida que las consultas son procesadas algunos agentes pueden ser incapaces de responder (y es como si se hubiesen ido del sistema, mientras que otros pueden ingresar a él (Figura 4)). La respuesta es retornada a Ω . Con las consultas subsiguientes, Ω y los otros agentes van a determinar qué agentes son sus favoritos (Figura 5). Los nuevos agentes que aparezcan como favoritos de Ω y de los otros agentes van ahora a responder las consultas (Figura 6). Este proceso se repite indefinidamente durante la operación del sistema. Nótese que los agentes pueden salir y entrar del sistema en cualquier momento.

La salida del sistema por no responder más (por ejemplo debido a escasez de recursos, caída brusca de la calidad, etc.) puede considerarse como lógica. La idea de partida lógica es que el agente puede continuar estando físicamente conectado al resto del sistema, pero desaparece para sus pares. Cuando un agente parte lógicamente solamente deja de responder mensajes. En cuanto consiga recursos podrá volver al sistema. Es una situación diferente respecto a cuando el agente se desconecta y decide no integrar más el sistema, lo que podría considerarse una partida física.

Los nuevos agentes pueden entrar al sistema en cualquier momento: no se requiere ninguna operación especial para que un agente deje el sistema lógicamente. La dificultad o costo de comunicación no es tenido en cuenta: todas son del mismo valor y no hay noción de proximidad topológica, es decir, no hay cosas tales como agentes cercanos o lejanos.

Figura 3: Dinámica del sistema - Paso 1. *(La consulta llega a todos los peers)*

Figura 4: Dinámica del sistema - Paso 2. *(Luego de un cierto número de interacciones cada agente a quien vale la pena consultar en cada caso).*

Figura 5: Dinámica del sistema - Paso 3. *(La partida y llegada de peers acontece con el tiempo)*

Figura 6: Dinámica del sistema - Paso 4. *(El sistema alcanza de nuevo un estado de equilibrio)*

Se supone que los agentes están escuchando permanentemente los mensajes, aunque no puedan responderlos, siempre que tengan los recursos adecuados para ello (típicamente, batería).

Basándose en el modelo diseñado, el sistema puede pasar de un modo sin restricciones a un modo operativo cuando sus agentes detectan ciertas condiciones tales como un número de mensajes no contestados muy alto o cuando se recibe una señal de limitación en el número de mensajes a usar. Un sistema como este tiene un amplio espectro de dominios de aplicación, desde herramientas de recomendación móviles a aplicaciones de toma de decisiones en entornos críticos. Es más, se provee evidencia empírica del uso del nivel 4 del modelo JDL para la gestión del proceso de IF (Blasch *et al.*, 2013; Steinberg *et al.*, 1999), ya que el modelo propuesto elige adaptativamente los datos a ser provistos eligiendo los datos de mejor calidad de las fuentes que responden (o sea, de las disponibles), mientras que las fuentes consultadas cambian con el tiempo. Esta es una contribución a la escasa investigación existente sobre este tópico (Solaiman *et al.*, 2015).

Se considera que los agentes del sistema pasan por cuatro etapas o fases durante su operación:

- **FASE 0:** Entrenamiento de los mecanismos de predicción/estimación (por ejemplo, redes neuronales).
- **FASE 1:** Interacción entre los agentes de acuerdo a las reglas mencionadas antes (por ejemplo responder si la calidad del agente es solo ocasionalmente menor que la del agente que lo consulta). Esta fase incluye el tiempo de estabilización, es decir, cuando cada agente determina cuales son los mejores agentes a ser consultados para cada campo. La FASE 2 comienza cuando se han determinado esos agentes.
- **FASE 2:** Operación de los agentes en el sistema.

Todos los agentes van a estar, en algún momento, en la misma fase. Cuando esto sucede, se puede decir que el sistema está en esa fase. Por lo tanto, el sistema pasa de una fase a otra, con fases intermedias híbridas, que son en las que la reorganización de los agentes sucede.

La evolución del sistema en el tiempo implica que los diferentes agentes pasan periódicamente por la FASE 1, cuando los mejores agentes deberían ser determinados de nuevo debido a los *peers* que perdieron la posibilidad de comunicarse. Como ya se dijo, cada vez más las redes son *peer-to-peer* a la vez que las comunicaciones hombre-máquina son más comunes. Las redes P2P pueden estar sujetas a limitaciones en sus recursos por distintas razones. Por esta razón sus miembros deberían tener la inteligencia suficiente como para optimizar la calidad de los resultados de salida usando sus recursos disponibles.

Las figuras 7 - 9 muestran el comportamiento de los tres tipos de agentes que existen para el sistema: el agente que realiza la consulta y que es externo al sistema (Ω u OMEGA), los agentes del sistema y los agentes externos al sistema que desean incorporarse a él.

- **AGENTE DEL SISTEMA** (figura 7): Hay dos grandes tipos de eventos que lo afecta: llegada de un mensaje y que se produzca el TO de una consulta que él realizó. El mensaje puede ser una consulta, una respuesta o una solicitud de ingreso a la red (CHECK-IN). Si es una consulta, decidirá por qué sub-campos consultar, y si lo hace, actualizará su inventario de recursos y arrancará el control del TO para esas consultas. Si en cambio es una respuesta verá si respondieron todos los consultados y si lo hicieron hará la fusión pertinente. Se devolverá la respuesta si la calidad obtenida es mayor que la del consultante y se sortea en caso de que sea menor. En el caso en que el modo sea simple, siempre se responde. Luego de responder se actualizan los inventarios (*stocks*) de recursos. Cuando es una solicitud de ingreso se analiza si hay recursos disponibles y si los hay se responde al solicitante. Cuando es un TO se lo contabiliza y la respuesta que se estaba esperando se toma vacía. Cuando se llega a cierta cantidad de TOs el agente pasa a trabajar de modo inteligente.

- **AGENTE EXTERNO** (figura 8): Lo que hará es elegir agentes al azar de la lista de bootstraps disponibles. Mandará mensajes de check-in a ellos hasta que alguien le conteste o se quede sin recursos.

- **OMEGA** (figura 9): OMEGA envía consultas a una serie de agentes favoritos (los que le han respondido con mayor calidad hasta el momento, o a “muchos” cuando es la primera vez que interactúa con el sistema). cuantos es “muchos” depende de la cantidad inicial de agentes del sistema y se supone que *muchos* $\ll$ *cantidad inicial de agentes*

en el sistema. Cuando recibe una respuesta ve si llegaron todas las respuestas para la consulta. si llegaron, realiza la fusión de las respuestas y podría realizar otra consulta. Permanentemente controla el momento en que se produce *time-out*, y es entonces que realiza la fusión con las respuestas recibidas hasta el momento. Asimismo con cada respuesta recibida actualiza su lista de favoritos.

5.4. Parámetros del modelo

Existe una serie de parámetros que deberían ser ajustados, tomando en cuenta el número de agentes del sistema, las velocidades de las redes de comunicación, la velocidad de proceso de los agentes, la complejidad de predecir cada campo, etc. Dichos parámetros son:

- TIEMPO.DE.TIME.OUT: Tiempo máximo de espera por una respuesta (el mismo para todos los agentes).
- MENS.PREVIO: Cuántos mensajes se intercambian antes de elegir el agente preferido en la predicción de un cierto campo.
- NRO.AG.A.CONSULTAR $\equiv \hat{N}$: Número máximo de agentes a consultar por cualquier agente (cuantos agentes se toma del tope de la lista de mejores agentes).
- MAX.NRO.MENSAJES.AG: Número máximo de mensajes que tienen los agentes al comenzar a actuar. Si es < 0 no se toma en cuenta.
- PESO.NIVEL $\equiv Z$: Entero no negativo que indica la importancia que se le da a tener una buena calidad en unos pocos pasos de consulta. Si es 0, el nivel no se toma en cuenta.
- UMBRAL.FRONTERA: Número de mensajes mínimo que tiene que tener un agente para contestar un CHECK-IN.
- LIMITES.TIME.OUTS: Cantidad de TOs que debe aceptar un agente antes de empezar a trabajar en modo inteligente.
- W: Matriz con los pesos relativos de los campos al calcular la calidad del registro m (donde la fila i son los pesos de los subcampos de m_i siendo m formado por $m_1, m_2, \dots, m_i, \dots$).

Figura 8: Diagrama de flujo del comportamiento de un agente externo al sistema.

Figura 9: Diagrama de flujo del comportamiento del agente Ω .

Parámetro \ Perfil	Seguridad y vigilancia	Velocidad	Precisión
TIEMPO.DE.TIME.OUT	90s	30s	180s
MENS.PREVIO	2	5	10
NRO.AGENTES	10	20	30 o más
NRO.AG.A .CONSULTAR	2	4	8 o más
MAX.NRO.MENSAJES.AG	70	25	70
PESO.NIVEL = Z	2 o más	1	0
UMBRAL.FRONTIERA	30	30	10
LIMITES.TIME.OUTS	10	5	30
NRO.MAX.CONSULTAS	1	2	3 o más

Tabla 4: Parámetros del sistema por tipo de aplicación

- NRO.MAX.CONSULTAS: Número máximo de consultas pendientes de respuesta para un agente.
- LIMITE.ACTIVIDAD: Cantidad de TOs que recibe un agente de otro antes de considerarlo inactivo.
- NUM_CONS: Número máximo de agentes a consultar por Ω al comenzar a interactuar con el sistema.

Debido al conocimiento detallado del sistema que se requiere para poder ajustar estos parámetros, se pueden definir varios perfiles de uso, es decir, combinaciones de parámetros con ciertos valores adaptados a una aplicación con valores por defecto que se asumen adecuados. Por ejemplo, los perfiles podrían ser: aplicaciones relacionadas con la seguridad y vigilancia, aplicaciones donde lo crítico es la velocidad de respuesta, aplicaciones donde lo crítico es la precisión de la respuesta, teniendo cada uno los siguientes parámetros (asignados en forma totalmente heurística), tal como se muestran en la tabla 4.

5.5. Descripciones de los procesos y comportamientos

Básicamente los comportamientos de los agentes son disparados por eventos. Se describen los principales eventos que se manejan en el modelo.

- 1) **LLEGA CONSULTA A Y:** cuando un agente Y recibe una consulta desde un

agente Z por cierta parte del mensaje original (campo m) identificado por un indicador único de mensaje *id_mensaje* lo que va a hacer es decidir si consulta o no a otros *peers* en forma aleatoria por algunos sub-campos de m elegidos al azar (si los hay) considerando para ello la calidad que tiene Y para ese sub-campo de m : cuanto menor calidad tenga mayor es la probabilidad de que consulte. Y lleva un control de que no envíe demasiadas consultas al mismo agente, para evitar la sobrecarga. Si Y no consulta por ningún sub-campo de m , Y contestará a Z con su propia estimación de m . Esto se describe en el pseudocódigo S1.

```

1: S1 EVENTO: LLEGA CONSULTA A Y
2: Descripción: Llega una consulta a Y desde Z por el campo  $m$  para id.mensaje
   .
3: _____
4:
5: Comienzo
6:     Para algunos campos compuestos de  $m$  elegidos al azar (incluyendo al propio  $m$ )
       hacer
7:         Consultar con probabilidad inversamente proporcional a la calidad de
8:         la aproximación de Y al subcampo a otros  $N$  agentes distintos de Z
9:         //Se controla que no se tengan demasiadas consultas a la vez para el mismo agente
10:    Fin para
11:    Si no hay consultas generadas por Y pendientes para subcampos de  $m$  y ese id.mensaje
       entonces
12:        Responder a Z por  $m$  y para id.mensaje
13:    Fin si
14:    Fin para
15: Fin
16: _____

```

2) **LLEGA RESPUESTA A Y:** cuando un agente Y recibe una respuesta desde un agente Z por cierta parte del mensaje original (campo m) identificado por un indicador único de mensaje *id_mensaje* lo que va a hacer es: primero, ver si el mensaje llegó antes de expirar el tiempo disponible (o sea si no fue un *time-out*), si no lo fue y si ya todos los agentes consultados por Y para ese campo y mensaje respondieron, se responderá a todos los agentes que habían consultado a Y por ese campo y ese mensaje. Si fuera un *time-out*, se descarta la información que proveía la respuesta, se lo contabiliza y si se excedió un número umbral de *time-outs* Y pasa a trabajar en modo inteligente, si ya no se estaba en él. Esto se describe en el pseudocódigo S2.

```

1: S2 EVENTO: LLEGA RESPUESTA A Y
2: Descripción: Llega una respuesta a Y desde Z por el campo  $m_t$  para
   id.mensaje
3: _____
4:
5: Comienzo
6:     Si la respuesta llegó antes del TIME OUT entonces
7:         Si se esta trabajando en modo inteligente entonces
8:             Contar el nro. de respuestas obtenidas para  $m_t$ 
9:         Fin si
10:    sino
11:        Contar el TIME OUT
12:        Si se excedió el umbral máximo entonces

```

```

13:           Pasar a modo inteligente ese agente Y
14:       Fin si
15:       Hacer respuesta=NULL
16:   Fin si
17:   Si todos los agentes ya respondieron para ese  $m_t$  e  $id.mensaje$  o si dio TIME OUT
   entonces
18:       Responder
19:   Fin si
20: Fin
21: _____

```

3) **LLEGA MENSAJE OK:** Cuando un agente X quiere ingresar a la red P2P envía mensajes de tipo CHECK-IN a una serie de agentes (que son conocidos públicamente) hasta que alguno Y le conteste OK, indicando que X ha sido incorporado como agente conocido por Y, es decir, que se lo incluyó en la red, o hasta que se le acaben los recursos a X. En ese momento X registra que ya no precisa enviar más mensajes de CHECK-IN cambiando un indicador que le dice si debe seguir intentando (la bandera de intentar conexión). Ver S3.

```

1: S3 EVENTO: LLEGA MENSAJE OK
2: Descripción: Llega un mensaje OK a X desde Y .
3: _____
4:
5: Comienzo
6:   Hacer Bandera de intentar conexión en X en 0
7: Fin
8: _____

```

4) **PROCEDIMIENTO: RESPONDER:** Cuando el agente Y ha recibido las respuestas para todas las consultas hechas para los sub-campos de m (o se produjo el *time-out*) calcula el valor de m haciendo la fusión del valor propio de Y con las respuestas recibidas y calcular la calidad del resultado. A su vez ese valor es enviado a todos los que consultaron a Y teniendo en cuenta que: si se está en modo inteligente la respuesta se devuelve siempre si la calidad de Y es mayor que la de quien lo consultó, pero si es menor hay una probabilidad p de que se responda con $p = (1 - Q) * MR$ siendo Q la calidad de Y y MR el porcentaje de mensajes restantes de Y, o sea, a medida que quedan menos mensajes a Y, para una misma calidad, es menos probable que responda aunque su calidad sea menor que la del consultante. Si no se está en modo inteligente Y responde siempre.

Recursivamente responden los agentes que tenían consultas pendientes por campos de los cuales m es sub-campo. Esos super-campos de m son llamados “padres” en el pseudocódigo S4.

```

1: S4 PROCEDIMIENTO RESPONDER
2: Descripción: Y responde formando el campo  $m$  que ahora puede ser cal-
   culado.
3: _____
4:
5: Comienzo
6:   Hacer fusión de respuestas y del valor calculado por Y //Si dio time-out la respuesta es
   el conjunto vacío
7:   Para cada padre de  $m_t$  que ahora puede ser calculado y que fue consultado por un Z
8:     Calcular valor  $V$  del padre
9:     Calcular la calidad  $Q$  de  $V$  según Y
10:    Si esta en modo inteligente entonces
11:      Si  $Q >$  calidad del consultante entonces
12:        Responder a Z con valor  $V$  y calidad  $Q$ 
13:      sino
14:        Con probabilidad proporcional a  $(1 - Q)MR$  Responder a Z con
15:        valor  $V$  y calidad //  $MR$  es el porcentaje de mensajes restantes,
16:         $MR = 1$  si no esta en modo inteligente
17:      Fin si
18:    Fin si
19:  Fin para
20: Fin
21: _____

```

5) **LLEGA UN CHECK-IN (“HOLA”)**: Cuando un agente X quiere ingresar a la red P2Penvia mensajes de tipo CHECK-IN a una serie de agentes (que son conocidos públicamente llamados *bootstrap*) hasta que alguno Y le conteste OK, indicando que X ha sido incorporado como agente conocido por Y, es decir, que se lo incluyó en la red o hasta que X quede sin recursos. Y le contestará a X siempre y cuando tenga suficientes recursos para hacerlo (en el pseudocódigo S5 se consideran solamente el número de mensajes). Cada agente *bootstrap* lleva una lista de agentes conocidos por él (Y incluiría en esa lista a X) y también en una lista de agentes activos para Y (los que pueden ser consultados).

```

1: S5 EVENTO: LLEGA UN CHECK-IN (HOLA)
2: _____
3:
4: Comienzo
5:   Si Y tiene mensajes restantes  $>$  umbral entonces
6:     Insertar X en estructuras de Y y en lista de ingresados en el sistema según Y

```

```

7:         Insertar X en la lista de agentes activos para Y (a quien Y puede llegar a con-
          sultar)
8:     Fin si
9: Fin
10: _____

```

6) **PROCESO: INTENTO DE CONEXIÓN:** Cuando un agente X desea incorporarse a la red P2P elige al azar agentes *bootstrap* y les envía una solicitud de conexión (mensaje de CHECK-IN), dejando pasar un tiempo aleatorio entre los mensajes, hasta que alguno conteste. Ver S6.

```

1: S6 PROCESO: INTENTO DE CONEXIÓN
2: Descripción: El agente X empieza a tratar de ingresar a la red (sistema)
3:
4: Comienzo
5:     Hacer bandera-conexión = 1 // Prende la bandera que indica que debe seguir mandando
        HOLA
6:     a agentes
7:     Mientras bandera-conexión = 1 repetir // mientras esa bandera esta prendida manda
8:         mensajes HOLA a los agentes al azar
9:         Seleccionar un agente al azar de la lista de bootstraps // Existe una lista de
        agentes
10:         que se anotan como voluntarios para ser bootstraps, y esa lista es pública. La incorporación
11:         a esa lista es un evento no descripto aquí.
12:         Mandar mensaje de CHECK-IN (HOLA) al agente elegido
13:         Esperar un tiempo aleatorio
14:     Fin repetir
15: Fin
16: _____

```

CAPÍTULO 6

TRABAJO EXPERIMENTAL

En este capítulo se explica el aporte A1 (ver sección 1.3), es decir, cómo el modelo inteligente presentado aumenta el rendimiento de la red.

6.1. Descripción general de los experimentos

Se realizaron estudios cuantitativos, asociados a la RQ1, para tres casos de estudio; la RQ2 fue respondida según se describe en el capítulo 7.

Dado que el número de agentes a considerar era en algunos casos muy grande (varios cientos) y que por cada mensaje que recibe un agente se generaría una hebra (*thread*) de procesamiento, lo cual requeriría mucha más capacidad de cómputo que la disponible en computadoras de escritorio como las que se disponían, en lugar de implementar directamente un sistema multi-agente se construyó un prototipo que simula las comunicaciones entre los agentes y el procesamiento de información asociado. Es decir, el prototipo reproduce el comportamiento de un sistema *peer-to-peer* incluyendo un tiempo aleatorio de respuesta de un agente. Este enfoque, además de permitir la realización de los experimentos, no requiere la utilización de ninguna plataforma de *software* específica, permite variar casi arbitrariamente las cantidades de agentes y facilita la recolección de datos.

Se realizaron experimentos sobre tres casos de estudio: el llamado A, asociado a la determinación de si un día es de alta cantidad de ozono; el B, relativo a clasificación de un automóvil como adecuado o no para ser comprado según sus características (precio,

confort, seguridad, etc.); y el C, relacionado con la clasificación de sitios web como de *phishing* o no.

En el caso A, los valores de $H(x)$ (la medida de la vaguedad de un ítem) son constantes para los distintos pares de ejecuciones (inteligente, simple), aunque distintos entre sí, para permitir totalizar por cantidad de mensajes para los distintos números de agentes. En los casos B y C la función $H(x)$ se tomó directamente igual a 0, es decir, los datos fueron considerados como *crisp* (o sea no vagos) y la medida de la incertidumbre, $u(x)$, fue tomada como el porcentaje de error de un agente al predecir el valor del campo x en los tres casos.

Toda la actividad de los agentes se registró en archivos de texto ASCII y luego analizada usando programas *ad hoc* escritos en Python 3.5 para los casos B y C. Algunos de dichos archivos resultaron bastante extensos, por ejemplo, 208000 líneas. Para el A se realizó directamente la recolección de datos durante la ejecución del sistema, evitándose así algunos análisis muy complejos de logs.

Las simulaciones fueron ejecutadas en computadores personales convencionales (PCs) ejecutando Microsoft Windows 7 y 10 Pro. Las ejecuciones fueron hechas en paralelo en 4 computadores con el propósito de acelerar el proceso de recolección de datos.

Se utilizó software de oficina estándar (Microsoft Excel 2016) para procesar los datos. Las pruebas estadísticas se realizaron con SPSS y Statistica.

Por convención se dice que el sistema está trabajando en modo inteligente cuando los *peers* se comunican usando el modelo propuesto aquí y en modo simple en otro caso, es decir, cuando los *peers* consultan otros agentes sin considerar el comportamiento pasado de los demás.

Cada ejecución para una terna (número de agentes, número de mensajes, modo) fue repetida al menos 32 veces (siendo el modo “simple” o “inteligente” como se explicó antes), calculándose los valores promedios para este conjunto de ejecuciones de forma de asegurar que los resultados fueran estadísticamente válidos. Cada ejecución fue independiente. El mismo conjunto de 30 mensajes enviados por Ω (para cada caso de estudio) fue usado para todas las ejecuciones.

El valor 30 se eligió porque clásicamente se lo considera un número “grande” al hablar de tamaños muestrales, como se verá más adelante. Debido a ello las ejecuciones, número

de consultas de Ω , etc. suelen ser 30 al menos, y en general 32.

Esta versión del prototipo no cambiaba automáticamente de modo simple a inteligente cuando se detectaba un número alto de *time-outs* (TOs). Gracias a ello se pudieron comparar las propiedades de los dos modos más claramente ya que de esa forma no habría ningún estado transitorio del sistema (cuando solo algunos agentes trabajan en modo inteligente) que pudiese oscurecer la comparación (por ejemplo, los resultados de dicha comparación podrían variar dependiendo por ejemplo cuántos agentes están trabajando en modo inteligente).

El número máximo de búsquedas en paralelo fue mantenido constante en 5. El tiempo de *time-out* (TO) fue de 100 segundos para cada agente.

Se utilizó la distribución de probabilidad uniforme cada vez que se tuvieron que considerar valores al azar.

Salvo que se especifique lo contrario, todos los demás parámetros fueron constantes a lo largo de los experimentos.

Los datos usados provinieron de tres juegos de datos tomados del repositorio en línea de la Universidad de California en Irvine, Estados Unidos (UCI) (Lichman, 2013), que se describirán en detalle más adelante.

6.2. Medida del rendimiento

Teniendo en cuenta la RQ1 definida en la sección 3, se usaron las siguientes métricas para la evaluación de procesos:

- 1) Tasa de éxitos: un éxito en el contexto de redes P2P se produce cuando se encuentra un recurso para satisfacer una demanda. En este caso será que logre obtener una respuesta con calidad mayor o igual que la que tenía inicialmente Ω . Un éxito es estricto cuando la calidad de la respuesta es estrictamente mayor que la de Ω y amplio en otro caso. Los éxitos estrictos son especialmente interesantes ya que son los que corresponden a los casos en los que el sistema fue realmente útil. En los resultados mostrados se llama “éxito” a los éxitos estrictos, por brevedad.
- 2) Tráfico medio: es el promedio de mensajes usados por agente (lo que puede ser

asociado con el costo de la consulta), ya sea por mensaje recibido desde Ω o de los usados para dar una respuesta con una calidad mínima.

- 3) Número de *time-outs* (TOs): número de veces que el sistema no contesta a Ω a tiempo.

El fin último de los experimentos realizados es el estudio de los siguientes tópicos, comparando en ambos casos el resultado de aplicar el modelo aquí descripto (inteligente) con un enfoque convencional (simple):

- Efecto del número de agentes y mensajes disponibles en el sistema sobre la tasa de éxitos.
- Efecto del número de agentes y mensajes disponibles sobre los TOs.

Para los casos B y C se consideró la RQ1 como formada por dos sub-preguntas: a) si el modelo inteligente requiere menos mensajes para que Ω obtenga una respuesta de calidad dada y b) si el modelo inteligente tiene una mayor tasa de éxitos.

6.3. Generalidades del análisis estadístico realizado

Dado que todos los experimentos se refieren a las medidas del comportamiento para un sistema operando en dos modos distintos y debido a que las ejecuciones, para cada conjunto de parámetros, son mutuamente independientes, los datos fueron analizados como muestras en pares (*paired samples*).

Cada par de muestras corresponde a 32 ejecuciones de una serie de 30 o más consultas que se repiten en los distintos experimentos (o sea que se analizan 30 o más pares).

Las hipótesis estadísticas fueron planteadas de forma que la hipótesis alternativa es el enunciado que se quiere realmente probar (asegurando así un soporte fuerte a la prueba, como lo explican Witte y Witte (Witte & Witte, 2013). Por ejemplo, si se quiere probar si la media de una medida en particular (tal como el número de éxitos), x , es mayor en el modo inteligente que en el simple la hipótesis nula planteada es $x_{INTELIG} \geq x_{SIMPLE}$. En otros casos, tales como el número de TOs o número de mensajes usados para alcanzar una calidad dada de la respuesta a Ω , $x_{INTELIG} \leq x_{SIMPLE}$.

Todas las pruebas de hipótesis se realizaron mediante una prueba de Student (t -test) de muestras apareadas. Previamente se aplicó la prueba de Shapiro-Wilk para verificar la normalidad de las diferencias entre las dos series ya que se considera la prueba más poderosa para ello (Chen, 1971; Raol, 2015). Cuando las diferencias no son normales hay varias posibilidades. Primero, dado que el tamaño de la muestra es grande (como $n > 30$ la muestra se considera grande de acuerdo a la literatura (Brayner *et al.*, 2014; Starnes *et al.*, 2014), se puede realizar directamente el t -test. Segundo, se pueden utilizar métodos no paramétricos como la prueba de Wilcoxon para verificar que las muestras que forman el par provienen de distintas distribuciones de probabilidad (es decir, que el modo inteligente tiene un efecto en el comportamiento del sistema (Witte & Witte, 2013)). Como otro punto que interesa esclarecer es cuál media es menor, se pueden determinar intervalos de confianza de las medias usando *bootstrapping*. Suponiendo que los intervalos son (a, b) y (c, d) entonces si $b < c$ se puede deducir que una media es menor que otra para un nivel de significación α . Esto se denota aquí como $\mathcal{I}(\mu_0) \prec \mathcal{I}(\mu_1)$ en las tablas 7 y 11, significando que el intervalo de confianza a nivel 0.05 para μ_0 es (a, b) y para μ_1 es (c, d) y que se cumple $b < c$. Tercero, se puede aplicar la transformación de Box-Cox (Sakia, 1992) a las dos series que forman los pares usando el mismo λ . Se calculan entonces las diferencias entre las transformadas y se aplica el t -test a las diferencias resultantes (Doksum & Wong, 1983). La aplicación de la transformación no asegura que se alcance la normalidad. Cuarto, se podrían analizar los datos y ver los gráficos Q-Q (Berkman & Reise, 2012) de forma de encontrar si algunos *outliers* están causando la no-normalidad. Finalmente también se realizó la prueba de Hodges-Lehmann (Hershberger, 2011) en varios casos para verificar que las diferencias entre las distribuciones (y respectivas medias y medianas) eran significativas. Estas pruebas son una forma de confirmar los resultados de aplicar el t -test si la condición de normalidad de las diferencias no se cumple. En los casos B y C, se usó la transformación de Box-Cox cuando las diferencias no fueron normales. Si esta transformación no conducía a la normalidad, se analizó los datos buscando *outliers*. Si los outliers encontrados eran menos del 10 % de la muestra se los removía, de la misma forma, se realizaron la prueba de signos de Wilcoxon y el *signed-rank test* para ver si las distribuciones de las dos series eran las mismas. También se determinaron intervalos de confianza de las medias con *bootstrapping* y se verificó si $\mathcal{I}(\mu_0) \prec \mathcal{I}(\mu_1)$ o $\mathcal{I}(\mu_0) \succ \mathcal{I}(\mu_1)$.

El nivel de significación mínimo usado en todas las pruebas fue $\alpha = 0,05$, aunque los niveles $\alpha = 0,01$ e incluso $\alpha = 0,001$ fueron empleados para algunas pruebas debido a las implicaciones de aceptar la hipótesis en la verificación de las bondades del modelo inteligente.

6.4. Caso A: Sensores (*Soft sensors*)

6.4.1. Generalidades

La concentración de ozono a nivel del suelo depende de varios factores y se puede considerar de naturaleza estocástica. Un nivel de ozono por sobre cierto umbral conocido es peligroso para la salud humana y afecta actividades como trabajos agrícolas, turismo, etc. Por lo tanto es necesario tener un sistema preciso de predicción del ozono para enviar las alertas necesarias antes de que alcance niveles peligrosos.

Una forma de realizar esas predicciones es usando modelos que incluyen fórmulas, parámetros y las leyes que gobiernan las reacciones físicas y químicas involucradas (Lichman, 2013). Sin embargo, debido al conocimiento limitado y al costo que tiene construir esos modelos, se puede intentar usar un *soft sensor*.

Los datos disponibles consisten en registros de 72 columnas, siendo cada una de ellas un parámetro (correspondiente a un campo en nuestro modelo) que determinan si el nivel de ozono es alto o no, por ejemplo temperatura máxima, temperatura promedio, humedad relativa, presión atmosférica a nivel del mar, precipitaciones, radiación total del día, temperaturas y vientos medidos a distintas horas del día. Algunos de esos parámetros pueden considerarse compuestos por otros de la lista (por ejemplo TT depende de T850, T500, Td850 y T500, KI depende de T850, T500, DP850 y DP700). El campo de decisión de si es un día de ozono alto o no se considera dependiente de todos los demás.

Para más detalles sobre los datos usados ver el Anexo A.

6.4.2. Resultados obtenidos

Se estudió el número medio de mensajes requeridos para contestar un conjunto de consultas (*queries*) provenientes de Ω teniendo en cuenta que se quiere esclarecer la RQ1

considerada como: a) si al usar el modelo inteligente se requiere menos mensajes que para conseguir una respuesta de cierta calidad y b) si el modelo inteligente tiene una tasa de éxitos más alta. Los resultados se muestran en la tabla 5. Se realizó cada prueba para 10,20,30,50 100, 200 agentes usando 50,100,200,500 mensajes.

Las figuras 10-13 muestran gráficamente los resultados. Notar el comportamiento del sistema al aumentar la cantidad de agentes o el número de mensajes disponibles. En esas gráficas se muestra el cociente de los valores Simple/Inteligente. Las tendencias lineales se grafican en línea punteada. La Figura 10 y la 12 (consumo, o sea, cantidad de mensajes por éxito estricto, en el eje y versus cantidad de agentes o mensajes en el eje x) señalan que el modo inteligente presenta ventajas sobre el simple al aumentar la cantidad de agentes o de mensajes disponibles por tener la tendencia de dicho cociente pendiente positiva. De la misma forma las figuras 11 y 13 (promedio de TOs versus cantidad de agentes o de mensajes disponibles) muestran cómo los TOs se van aumentando al crecer los agentes o los mensajes disponibles en el sistema.

Asimismo la *research question* RQ1 está relacionada con el número de consultas que no son respondidas debido a los TOs (número medio de TOs dependiendo del número de mensajes y número medio de TOs dependiendo del número de agentes). Se observa que el modo inteligente reduce sistemáticamente el número de TOs (mensajes que no son contestados porque el tiempo predefinido para responder la consulta expira), promediado sobre los distintos números de mensajes disponibles inicialmente; simétricamente, la operación en modo inteligente siempre reduce la cantidad de TOs promediados sobre todos los agentes.

Por lo tanto los resultados fueron positivos, confirmando la RQ1 para este caso de estudio.

Figura 10: Consumo de mensajes en el sistema al aumentar los agentes.

Figura 11: TOs al aumentar los agentes.

Figura 12: Consumo de menssajes al aumentar los mensajes.

Figura 13: TOs del sistema al aumentar los mensajes.

6.4.3. Análisis estadístico

En la mayoría de los casos se rechazó la hipótesis nula y la hipótesis alternativa se confirmó (el modelo inteligente tuvo mejor desempeño que el método simple), como se muestra en la tabla 5.

Las veces en que la hipótesis nula no pudo ser rechazada en forma global (es decir, para todas las calidades o para todos los diferentes números de mensajes o agentes), como en el caso de las pruebas a) y b) de la tabla 5 fue para 10 agentes. Esto significa que el modelo no es adecuado para pequeñas cantidades (por ejemplo, menos de 20) de *peers*; de todos modos el mismo está pensado para redes P2P, las que seguramente cuentan con más de 20 agentes.

Hipótesis a probar	Resultados obtenidos
<p>a) La cantidad de veces que el sistema tarda demasiado en dar una respuesta o da una respuesta de calidad menor que la de Ω es menor en el inteligente que en el simple.</p> <p>μ_0 promedio de time outs generados en modo inteligente</p> <p>μ_1 ídem pero en modo simple</p> <p>$H_0 : \mu_0 \geq \mu_1$</p>	<p>Las diferencias son normales para 20,30,50, 100 y 200 agentes. El test de Student indica rechazar H_0 siempre. Para el caso de 10 ag el test de Wilcoxon indica distribuciones diferentes. La mediana de las diferencias muestrales es 27. El estimador de Hodges-Lehman es -27 para 10 agentes.</p>
<p>b) Dada una calidad α la cantidad promedio de mensajes usados en el sistema para obtener una respuesta de esa calidad al menos es menor en el modo inteligente que en el simple.</p> <p>μ_0 promedio de mensajes usados en modo inteligente para obtener una respuesta de calidad α al menos</p> <p>μ_1 ídem pero en modo simple</p> <p>$H_0 : \mu_0 \geq \mu_1$</p>	<p>Las diferencias entre las series de datos son normales (prueba de Shapiro-Wilk) para 10,20 y 30 ag. se rechaza H_0 para 20 y 30 ag, con una prueba de Student y a nivel 0.05. Para 50, 100 y 200 ag la prueba de Wilcoxon señala distribuciones distintas, siendo las medianas de las diferencias de las muestras negativas siempre. El estimador de Hodges-Lehman es 2586, 2171 y 1313 para 50, 100 y 200 agentes respectivamente.</p>
<p>c) El promedio de mensajes usados por éxito estricto es menor en el inteligente que en el simple. μ_0 promedio de mensajes usados en modo inteligente por éxito estricto (“rendimiento”)</p> <p>μ_1 ídem pero en modo simple</p> <p>$H_0 : \mu_0 \geq \mu_1$</p>	<p>Como las muestras son de más de 30 elementos, se podría ignorar las condiciones de normalidad y en ese caso un test de Student arroja rechazo de H_0 en todos los casos. Considerando la normalidad (prueba de Shapiro-Wilk), para 50 y 200 no son normales las diferencias, y en ese caso la prueba de Wilcoxon señala distribuciones diferentes, siendo las diferencias normales en el resto de los casos. El estimador de Hodges-Lehman es 3782 y 37858 para 50 y 200 agentes respectivamente.</p>

Tabla 5: Pruebas realizadas y resultados obtenidos para el caso A.

6.5. Caso B: Automóviles

6.5.1. Generalidades

En este caso, la meta es clasificar si un automóvil es o no aceptable (es decir, si es aconsejable comprarlo) basándose en un conjunto de atributos.

La cantidad de niveles de fusión de datos es mayor en este caso de estudio que para el C y hay muchos menos campos elementales en este que en el C; el campo resultante en el C es formado haciendo la fusión del doble de campos que en el B. En suma, los dos casos de estudio son de alguna forma complementarios. La información manejada dentro de este problema es estructurada como sigue:

CAR = Aceptabilidad del auto, compuesta por PRICE y TECH

- PRICE = Precio, compuesto por
 - Purchase price (precio de compra)
 - Maintenance price (precio de venta)
- TECH = Especificaciones técnicas, compuesto de COMFORT y Safety
 - COMFORT = confort, compuesto de
 - Number of doors (número de puertas)
 - Number of seats (número de pasajeros)
 - Size of boot (tamaño del maletero)
 - Safety (seguridad)

Para más detalles sobre los datos usados ver el Anexo B.

6.5.2. Resultados obtenidos

Las pruebas realizadas están en la tabla 6 y sus resultados son mostrados en las tablas 7 y 8.

La figura 14 (número promedio de éxitos en el eje y versus mínimo número de mensajes disponibles en el eje x) y la figura 19 (promedio de mensajes necesarios para obtener una

calidad dada para 30 agentes en el eje y vs. las distintas calidades en el eje x) muestran gráficamente parte de los resultados. Concretamente se estudió:

- El promedio de éxitos estrictos para un número de mensajes dados (figura 14).
- El promedio de mensajes requeridos para alcanzar una calidad dada mínima en la respuesta a Ω (figura 15).
- La cantidad promedio de TOs que se producen para una cantidad de mensajes inicial dada (figura 16).
- El número promedio de éxitos para un número dado de agentes (figura 17).
- El promedio de TOs para un número de agentes (figura 18).
- El promedio de mensajes necesarios para obtener una calidad dada para 30 agentes (figura 19).
- La frecuencia con que se da una calidad mayor o igual a α , para un α dado (figura 20).

Si bien las pruebas muestran que aunque el sistema da resultados que son éxitos estrictos más a menudo cuando trabaja en el modo simple que en el inteligente, esto es cierto solo para las menores calidades, mientras que el sistema devuelve más a menudo un éxito estricto para la mayor calidad (1.0) si está trabajando en modo inteligente. Ver la tabla 7 y la figura 14 (número promedio de éxitos estrictos en el eje y para un número de mensajes mínimo en el eje x).

Además:

a) Con respecto al número de consultas que no son respondidas debidos a los TOs (el promedio de TOs dependiendo del número de mensajes y el promedio de TOs dependiendo del número de agentes se pueden ver en las figura 16 y 18 respectivamente): se observa que el modo inteligente reduce el número de TOs al considerarlo promediado los distintos números de mensajes disponibles inicialmente, simétricamente, la operación en modo inteligente siempre reduce el número de TOs promediado sobre todos los números de agentes.

b) En cuanto al número de éxitos por agente (promedio de éxitos estrictos para un número máximo de agentes y para un número de mensajes inicialmente disponibles, figuras 14 y 17 respectivamente). Se encontró que el número de éxitos estrictos que pueden alcanzarse al cambiar el número de agentes en el caso B es siempre mayor en el modo inteligente que en el simple). Esto es complementado por los resultados del caso C donde el modo inteligente tiene mejor rendimiento que el modo simple mientras haya 30 o menos agentes. La figura 17, donde el eje y denota el promedio de éxitos estrictos y el eje x el número original de mensajes por agente) muestra parte de lo encontrado. Se halló que el número de éxitos estrictos que se alcanzan al cambiar el número de agentes es siempre mayor en el modo inteligente que en el simple.

Por lo tanto los resultados son también positivos para la RQ1 en este caso de estudio. La sección 6.5.3 reporta con detalle las pruebas estadísticas que confirman que la estrategia inteligente propuesta tiene un efecto estadísticamente significativo en los resultados (principalmente, el número de mensajes requeridos y los éxitos logrados).

Figura 14: Promedio de éxitos estrictos para un número mínimo dado de mensajes.

Figura 15: Número promedio de mensajes requeridos para alcanzar una calidad mínima dada (todos los agentes).

Figura 16: Número promedio de TOs para una cantidad de mensajes dada.

Figura 17: Promedio de éxitos para un número de agentes mínimo.

Figura 18: Promedio de TOs para distintas cantidades de agentes.

Figura 19: Promedio de mensajes necesarios para obtener una calidad dada para 30 agentes.

Figura 20: Frecuencia con que se obtiene una calidad mínima.

6.5.3. Análisis estadístico

Se describen en la tabla 6 las distintas pruebas estadísticas que se realizaron, indicando las hipótesis a probar. Los resultados de las pruebas estadísticas y el tamaño muestral que las respalda se resumen en la tabla 7. La tabla 8 analiza los casos en que la hipótesis nula es rechazada parcialmente (es decir, sólo para algunas combinaciones de número de agentes y de mensajes).

Hipótesis	Medias	Prueba
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para una cantidad mínima de mensajes γ , con γ dado μ_1 Ídem μ_0 , pero en modo simple	1
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para una cantidad dada de mensajes iniciales μ_1 Ídem μ_0 , pero en modo simple	2
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para una cantidad máxima de agentes δ , con δ dado μ_1 Ídem μ_0 , pero en modo simple	3
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para un nro. de agentes dado μ_1 Ídem μ_0 , pero en modo simple	4
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para una cantidad mínima de mensajes γ , con γ dado μ_1 Ídem μ_0 , pero en modo simple	5
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para un número de mensajes dados μ_1 Ídem μ_0 , pero en modo simple	6
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para una cantidad máxima de agentes δ , con δ dado μ_1 Ídem μ_0 , pero en modo simple	7

Tabla 6: Pruebas a realizar para el Caso B.

Hipótesis	Medias	Prueba
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para una cantidad maxima de agentes δ , con δ dado μ_1 Ídem μ_0 , pero en modo simple	7
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para una cantidad de agentes dados μ_1 Ídem μ_0 , pero en modo simple	8
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 promedio(total éxitos en la iteación x /total de mensajes usados en la iteración x) para todos los nros. de agentes y nros de mensajes iniciales. μ_1 Ídem μ_0 , per\begin{sidewaystable}o en modo simple	9
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de mensajes, en modo inteligente, necesarios para obtener una calidad mínima β , con β dado μ_1 Ídem μ_0 , pero en modo simple	10
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de mensajes, en modo inteligente necesarios para obtener una calidad dada μ_1 Ídem μ_0 , pero en modo simple	11
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Nro. de veces que se respondió con una calidad mayor o igual a α dada en modo inteligente μ_1 Ídem μ_0 , pero en modo simple	12

Tabla 6. Pruebas a realizar para el Caso B.(*continuación*)

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
1	Se rechaza H_0 para $\alpha = 0.05$ y todas las cantidades de mensajes	Todas son normales salvo para 1000 mensajes. Para 1000 el test de Wilcoxon signed dice que son distintas distribuciones. $\mathfrak{S}(\mu_0) \not\prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es -6.7	32
2	Se rechaza H_0	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es 31.167 a nivel 0.05.	257
3	Para $\delta \in \{5, 10, 15, 20, 30, 100, 300, 600\}$: Se rechaza H_0 en todos los casos (para todo δ) y para $\alpha \in \{0.01, 0.05\}$	Todas las series de diferencias son normales a nivel 0.05.	32
4	Se rechaza H_0 para $\alpha = 0.05$	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es 26.25 a nivel 0.05	256

Tabla 7: Resultados de las pruebas para el Caso B

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
5	Para $\gamma \in \{5, 10, 15, 20, 30, \}$ 100, 1000} : Se rechaza H_0 en todos los casos (para todo γ) y para $\alpha \in \{0.01, 0.05\}$	Todas son normales	32
6	Se rechaza H_0 para $\alpha = 0.05$ Para las 164 primeras diferencias se rechaza H_0 para $\alpha \in \{0.001, 0.01, 0.05\}$	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \not\prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es -8.7 a nivel 0.05. Para la serie de las primeras 164 diferencias es normal	225
7	Para $\delta \in \{5, 10, 15, 20, 30, 100, 1000\}$: .Se rechaza H_0 en todos los casos (para todo δ), para $\alpha \in \{0.01, 0.05\}$	Todas las series de diferencias son normales a nivel 0.05.	32
8	Se rechaza H_0 para $\alpha = 0.05$	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \succ \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es -9.439 a nivel 0.05	257
9	Se rechaza H_0 a nivel $\alpha \in \{0.01, 0.05\}$	Las diferencias son normales a nivel 0.05	48

Tabla 7. Resultados de las pruebas para el Caso B (*continuación*).

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
10	Se rechaza H_0 para todas las calidades a nivel $\alpha \in \{0.01, 0.05\}$	Todas las diferencias son normales a nivel 0.05.	35
11	Se rechaza H_0	Dejando de lado la normalidad se podría rechazar H_0 hasta para 0.001 Las pruebas de signos y de Wilcoxon para muestras relacionadas señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es 501.91 a nivel 0.05.	334
12	Se rechaza H_0 a nivel $\alpha \in \{0.01, 0.05\}$ para $\alpha < 1$	Para las calidades que no son <0.1 ni 1.0 las diferencias son normales. Para <0.1 y 1.0 las pruebas de signos y de Wilcoxon para muestras relacionadas señalan 2 distribuciones distintas. Para 1.0 es $\mathfrak{S}(\mu_1) \prec \mathfrak{S}(\mu_0)$ pero para <0. es $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) para 1.0 es -1 a nivel 0.05 . O sea que el modelo simple tiende a dar mas respuestas de menor calidad que el inteligente (menos respuestas de calidad 1.0)	32

Tabla 7. Resultados de las pruebas para el Caso B (*continuación*).

Prueba	Resultado de la prueba	Significado
5	H_0 rechazada en todos los casos	$H_0 : \mu_0 \leq \mu_1$ con μ_0 : Cantidad de éxitos, en modo inteligente, para una cantidad mínima de mensajes γ , con γ dado Significado: Se obtienen más éxitos cuando hay pocos mensajes disponibles (cantidad mínima < 20).
7	H_0 rechazada en todos los casos	$H_0 : \mu_0 \leq \mu_1$ con μ_0 : Cantidad de éxitos, en modo inteligente, para una cantidad maxima de agentes δ , con δ dado Significado: Cuando hay pocos agentes (≤ 30) el modo inteligente produce más éxitos. Esto puede deberse al fenómeno de la “sabiduria de las multitudes” (Miller & Steyvers, 2011)
12	H_0 rechazada cuando $\alpha < 1$	$H_0 : \mu_0 \geq \mu_1$ con μ_0 : Nro. de veces que se respondió con una calidad mayor o igual a α , con α dado, en modo inteligente Significado: El modo inteligente tiende a aumentar la calidad de las respuestas, ya que produce más resultados de calidad ≥ 0.6 .

Tabla 8: Análisis de los resultados de las pruebas con rechazo parcial de H_0

6.5.4. Discusión de los resultados

El modelo inteligente arrojó mejores resultados que la versión simple en distintos aspectos, como se muestra gráficamente y es confirmado estadísticamente en la sección 6.5.3 y se resume a continuación. Primero, el promedio de mensajes requeridos para obtener una respuesta con una calidad mínima dada es menor. Esto fue demostrado por las pruebas 10 y 11 (tablas 6 y 7). Segundo, el número de éxitos estrictos para un número dado de mensajes iniciales fue mayor como lo muestran las pruebas 5 y 6 (tablas 6 y 7). Segundo, se observa a partir de la figura 14 que el número de éxitos estrictos en ambos casos aumenta con el número inicial de mensajes, aunque este resultado no fue validado estadísticamente. Tercero, el cociente del número de éxitos respecto al número de mensajes usados es mayor en el modo inteligente que en el modo simple. La prueba 9 de la tabla 7 confirma este punto ya que el número promedio de éxitos para una cantidad dada de mensajes es mayor en el modo inteligente que en el simple.

Respecto a los TOs, hay menos TOs respecto a los agentes y mensajes en el modo inteligente. El número de TOs es más o menos constante en el modo inteligente y casi constante en modo simple aunque en este caso se encontró que la tendencia tiene una pendiente positiva. Por otra parte, el modo inteligente conduce al sistema a causar menos TOs para cualquier número inicial de mensajes. Las tendencias fueron de la misma forma sistemáticamente mejores en el modo inteligente. Las pruebas 1, 2, 3 y 4 de las tablas 6 y 7 confirman esto.

Un aspecto importante es la frecuencia de las respuestas de alta calidad. En el modo inteligente, las respuestas de alta calidad son más frecuentes: observando la figura 20 se encuentra que las respuestas de alta calidad fueron ligeramente más frecuentes en el modo simple que en el inteligente. La prueba 12 de las tablas 6 y 7 explica el resultado. La calidad 1.0 es más frecuente en el modo inteligente que en el simple. Considerando el hecho que las frecuencias deben siempre sumar 1, se concluye que hay una mejora significativamente en la calidad de las respuestas en el modo inteligente ya que hay más respuestas de la mayor calidad (1.0).

6.6. Caso C: Sitios web

6.6.1. Generalidades

Se busca determinar si un sitio web (es decir, una dirección URL) es de *phishing* a partir de una serie de características de esa dirección y un conjunto de reglas que se aplican basadas en atributos del sitio. Los detalles están descritos en el Anexo C. Los registros contienen cuatro campos compuestos, los que a su vez dependen de los valores de otros campos que por brevedad no se mencionan. La clasificación del sitio (CLASS) depende de los cuatro campos compuestos:

- a) ADDRESS BAR BASED FEATURES
- b) ABNORMAL BASED FEATURES
- c) HTML- AND JAVASCRIPT-BASED FEATURES
- d) DOMAIN BASED FEATURES

Por ejemplo ADDRESS BAR BASED FEATURES depende de :

- Uso de la dirección IPs o URL larga para ocultar la parte sospechosa.
- Uso de servicios que abrevian la URL.
- URL conteniendo “@” y usando “//” .
- Adición de un prefijo o sufijo separado por - al dominio.
- Sub dominio y multiples sub-dominios.
- HTTPS (*Hyper Text Transfer Protocol with Secure Sockets Layer*).
- *Domain Registration Length*
- Favicon.
- Uso de puertos no estándares
- Existencia de la cadena “HTTPS” en la parte de nombre del dominio de la URL.

Para más información sobre los datos, ver el Anexo C.

6.6.2. Resultados obtenidos

Los resultados para este caso de estudio se muestran a continuación. El comportamiento del modelo propuesto mostró ser bueno también en este caso. Se realizaron los mismos análisis que para el caso B. Las tablas 9 y 11 muestran que la parte a) de la RQ1 se confirma de nuevo en este caso. La figura 21 ilustra que el modelo inteligente da calidades altas (de 0.7 o más) más a menudo que el modelo simple. Con respecto a la parte b) las figuras 22 y 23 confirman esta cuestión. En esta realización se encontró que el modelo inteligente es mejor cuando hay más de 100 agentes en el sistema (ver figura 24), aunque esto no es válido estadísticamente (ver tablas 9, 10 y 11), en el sentido de que la gráfica muestra una superioridad del modelo e-P2P para más agentes pero eso no se pudo confirmar estadísticamente. La figura 25 muestra cómo el modelo inteligente obtiene más éxitos estrictos que el simple para una misma cantidad de mensajes mínimo, y que esa tendencia aumenta con dicha cantidad.

Los gráficos conducen a la conclusión que la pregunta analizada también se cumple en el caso C si hay ciertas restricciones en el número de agentes.

La sección 6.6.3 reporta las pruebas estadísticas que confirman el significado de estos resultados.

Figura 21: Frecuencia promedio de una calidad mínima dada (todos los agentes).

Figura 22: Promedio de TOs para distintos números de mensajes.

Figura 23: Promedio de time-outs para distintos números de agentes.

Figura 24: Número promedio de éxitos estrictos para una cantidad máxima de agentes.

Figura 25: Promedio de éxitos estrictos.

6.6.3. Análisis estadístico

Al igual que en el caso B se busca realizar pruebas de hipótesis que permitan comprobar las bondades del modelo e-P2P ya sea en menor número de TOs como en un mejor rendimiento respecto del modelo “simple” P2P. La tabla 9 describe las distintas hipótesis a probar.

Hipótesis	Medias	Prueba
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para una cantidad mínima de mensajes γ , con γ dado μ_1 Ídem μ_0 , pero en modo simple	13
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs para un nro. inicial de mensajes dado μ_1 Ídem μ_0 , pero en modo simple	14
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para una cantidad maxima de agentes δ , con δ dado μ_1 Ídem μ_0 , pero en modo simple	15
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de time-outs, en modo inteligente, para una cantidad de agentes dada μ_1 Ídem μ_0 , pero en modo simple	16
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para una cantidad mínima de mensajes γ , con γ dado μ_1 Ídem μ_0 , pero en modo simple	17
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos para una cantidad de mensajes iniciales dada μ_1 Ídem μ_0 , pero en modo simple	18

Tabla 9: Pruebas a realizar para el Caso C.

Hipótesis	Medias	Prueba
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente, para una cantidad maxima de agentes δ , con δ dado μ_1 Ídem μ_0 , pero en modo simple	19
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 Cantidad de éxitos, en modo inteligente para una cantidad de agentes dada μ_1 Ídem μ_0 , pero en modo simple	20
$H_0 : \mu_0 \leq \mu_1$ $H_1 : \mu_0 > \mu_1$	μ_0 promedio(total éxitos en la iteación x /total de mensajes usados en la iteración x) para todos los nros. de agentes y nros de mensajes iniciales. μ_1 Ídem μ_0 , pero en modo simple	21
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de mensajes, en modo inteligente, necesarios para obtener una calidad mínima β , con β dado μ_1 Ídem μ_0 , pero en modo simple	22
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Cantidad de mensajes, en modo inteligente, necesarios para obtener una calidad dada μ_1 Ídem μ_0 , pero en modo simple	23
$H_0 : \mu_0 \geq \mu_1$ $H_1 : \mu_0 < \mu_1$	μ_0 Nro. de veces que se respondió con una calidad mayor o igual a α dada en modo inteligente μ_1 Ídem μ_0 , pero en modo simple	24

Tabla 9. Pruebas a realizar para el Caso C (*continuación*).

Prueba	Resultado	Significado
17	H_0 rechazada para $\gamma < 20$	<div> $H_0 : \mu_0 \leq \mu_1$ con μ_0: Cantidad de éxitos, en modo inteligente, para una cantidad mínima de mensajes γ, con γ dado </div> <p>Significado: Se obtienen más éxitos cuando hay pocos mensajes disponibles (cantidad mínima < 20).</p>
19	H_0 rechazada en los casos $\delta \leq 30$	<div> $H_0 : \mu_0 \leq \mu_1$ con μ_0: Cantidad de éxitos, en modo inteligente, para una cantidad maxima de agentes δ, con δ dado </div> <p>Significado: Cuando hay pocos agentes (≤ 30) el modo inteligente produce más éxitos. Esto puede deberse al fenómeno de la “sabiduria de las multitudes” (Miller & Steyvers, 2011)</p>
24	H_0 rechazada cuando $\alpha < 0.6$	<div> $H_0 : \mu_0 \geq \mu_1$ con μ_0: Nro. de veces que se respondió con una calidad mayor o igual a α, con α dado, en modo inteligente </div> <p>Significado: El modo inteligente tiende a aumentar la calidad de las respuestas, ya que produce más resultados de calidad ≥ 0.6.</p>

Tabla 10: Análisis de los resultados de las pruebas con rechazo parcial de H_0

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
13	Se rechaza $H_0 : \mu_0 \geq \mu_1$ para $\alpha = 0.05$ y todas las cantidades de mens $\gamma \in$ $\{5, 10, 15, 20,$ $30, 1000\}$	Ninguna es normal En todos los casos el test de Wilcoxon signed dice que son distintas distribuciones. En todos los casos es $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$.	32
14	Se rechaza $H_0 : \mu_0 \geq \mu_1$ $\alpha \in \{0.01, 0.05\}$	La serie de diferencias es normal.	257
15	Para $\delta \in$ $\{5, 10, 15, 20,$ $30, 100, 300, 600\}$: Se rechaza $H_0 : \mu_0 \geq \mu_1$ en todos los casos (para todo δ) y para $\alpha \in \{0.01, 0.05\}$	Las series de diferencias son normales a nivel 0.05 salvo para 15 y 20 ag. Para 15 y 20 La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. En los dos casos es $\mathfrak{S}(\mu_1) \succ \mathfrak{S}(\mu_0)$. Eliminando solo 3 outliers en cada caso ya se tiene una serie normal.	32
16	Se rechaza $H_0 : \mu_0 \geq \mu_1$ para $\alpha \in \{0.01, 0.05\}$	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es 11.35 a nivel 0.05 La transf de Box-Cox no es útil	256

Tabla 11: Resultados de las pruebas del caso C.

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
17	Para $\gamma \in \{5, 10, 15, 20, 30, 100, 1000\}$: Se rechaza $H_0 : \mu_0 \leq \mu_1$ en los casos $\gamma < 20$ y para $\alpha = 0.05$. O sea que a nivel 0.05, el inteligente obtiene más éxitos cuando hay pocos mensajes (cantidad mínima < 20)	Todas las diferencias son normales a nivel 0.05	32
18	Se rechaza $H_0 : \mu_0 \leq \mu_1$ para $\alpha \in \{0.01, 0.05\}$ Para las 164 primeras diferencias se rechaza H_0 para $\alpha \in \{0.001, 0.01, 0.05\}$	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \neq \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es -5.086 a nivel 0.05. Para la serie de las primeras 164 diferencias es normal y se puede aplicar el t-test La transf de Box-Cox no es útil	225

Tabla 11. Resultados de las pruebas del Caso C (*continuación*).

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
19	Para $\delta \in \{5, 10, 15, 20, 30, 100, 300, 600\}$: Se rechaza $H_0 : \mu_0 \leq \mu_1$ en los casos $\delta \leq 30$, para $\alpha \in \{0.01, 0.05\}$ O sea, el modo inteligente da más éxitos cuando hay no más de 30 agentes (≤ 30).	Todas las series de diferencias para $\gamma > 5$ son normales a nivel 0.05. Para $\gamma = 5$ las pruebas de signos y de Wilcoxon para muestras relacionadas señalan 2 distribuciones distintas. Es $\mathfrak{S}(\mu_1) \succ \mathfrak{S}(\mu_0)$. La mediana de las diferencias (test de Hodges-Lehman) es 16 a nivel 0.05	32
20	Se rechaza $H_0 : \mu_0 \leq \mu_1$ para $\alpha \in \{0.01, 0.05\}$	La serie de diferencias no aparece normal, debido tal vez a n muy grande. La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_1) \succ \mathfrak{S}(\mu_0)$. La mediana de las diferencias (test de Hodges-Lehman) es -5.75 a nivel 0.05 Se aplica la transformacion de Box-Cox a los datos originales y a ellos se le hace el t-test.	257
21	Se rechaza $H_0 : \mu_0 \leq \mu_1$ para $\alpha \in \{0.01, 0.05\}$	La serie es normal	53

Tabla 11. Resultados de las pruebas del Caso C (*continuación*).

Prueba	Resultado	Normalidad de las series de diferencias	Tamaño muestral (n)
22	Se rechaza $H_0 : \mu_0 \geq \mu_1$ para las calidades < 1.0 a nivel $\alpha \in \{0.05\}$ En modo inteligente se pueden llegar a usar tantos o más mensajes que en simple para alcanzar calidades $= 1.0$, para otras calidades se usan menos mensajes.	Todas las diferencias son normales a nivel 0.05 salvo para calidad ≥ 0.2 . La prueba de signos y la de signos relacionadas de Wilcoxon señalan 2 distribuciones distintas. $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) es 426.81 a nivel 0.05	35
23	Se rechaza $H_0 : \mu_0 \geq \mu_1$ $\alpha \in \{0.001, 0.01, 0.05\}$	Es normal a nivel 0.05	334
24	Se rechaza H_0 a nivel $\alpha=0.05$ para las calidades < 0.6	Para las calidades que no son < 0.1 ni 1.0 las diferencias son normales. Para < 0.1 y 1.0 las pruebas de signos y de Wilcoxon para muestras relacionadas señalan 2 distribuciones distintas. Para 1.0 es $\mathfrak{S}(\mu_1) \prec \mathfrak{S}(\mu_0)$ pero para $< 0.$ es $\mathfrak{S}(\mu_0) \prec \mathfrak{S}(\mu_1)$. La mediana de las diferencias (test de Hodges-Lehman) para 1.0 es -1 a nivel 0.05 . O sea que el modelo simple tiende a dar mas respuestas de menor calidad que el inteligente (menos respuestas de calidad 1.0)	32

Tabla 11. Resultados de las pruebas del Caso C (*continuación*).

6.6.4. Discusión de los resultados.

El modelo inteligente arrojó mejores resultados que la versión simple en distintos aspectos, como se muestra gráficamente en la sección 6.6.2 y se confirma estadísticamente en la sección 6.6.3. En efecto:

En cuanto al número promedio de mensajes requeridos para obtener una respuesta con una calidad mínima fue menor en el modelo inteligente, lo que quedó demostrado con las pruebas 22 y 23. Asimismo el número de éxitos estrictos para un número inicial de mensajes dado fue mayor en el modelo inteligente que en el común, tal como lo confirman las pruebas 17 y 18. Además, mirando la figura 25 se constata que el número de éxitos estrictos aumenta con la cantidad de mensajes disponibles inicialmente, al menos para el sistema estudiado (es decir, no se lo pudo probar estadísticamente). En cuanto a los TOs, éstos fueron menos tanto con respecto a los agentes y los mensajes. Si bien en este caso el número de TOs es más o menos constante en modo simple con distintos números de agentes, el modo inteligente lleva a que el sistema cause menos TOs para cualquier número de mensajes iniciales. Las tendencias son asimismo sistemáticamente mejores para el modo inteligente respecto al modo simple. Las pruebas 13, 14, 15 y 16 confirman esto. En cuanto a la calidad de las respuestas obtenidas, se observa que las respuestas de alta calidad son más frecuentes al utilizar el modo inteligente. Esto se desprende de la observación de la figura 21: se encuentra que las respuestas con calidades altas son ligeramente más frecuentes en el modo simple que en el modo simple. Las tendencias lineales de los dos modos de operación son casi idénticas en ambos casos. La prueba 24 explica los resultados. Las calidades 0.9 y 1.0 son más frecuentes en el modo inteligente que en el simple. Tomando en cuenta que las frecuencias deben siempre sumar 1, se concluye que hay una mejora significativa en la calidad de las respuestas en el modo inteligente, ya que hay más respuestas con las mayores calidades (0.9 y 1.0). Por último el cociente del número de éxitos sobre el número de mensajes usado es mayor en el modo inteligente que en el modo simple, lo que es confirmado por la prueba 21 de la tabla 9.

6.7. Conclusiones estadísticas generales

Para el caso A se pudo rechazar la hipótesis nula (o sea dar una respuesta positiva a la RQ1) en un 80 % de los casos y en el otro 20 % hay evidencias de que las distribuciones de probabilidad de los variables analizadas son distintas, es decir, que el modelo tiene una influencia positiva en el rendimiento del sistema.

Para los casos B y C, la mayoría de las veces ($21/24 \simeq 87,5\%$), la hipótesis nula fue rechazada y la hipótesis alternativa fue confirmada (el método inteligente propuesto tuvo mejor desempeño que el método simple), como se muestra en la tabla 7 y tabla 11 respectivamente. Siempre que la hipótesis nula no fue rechazada en conjunto (es decir, para todas las calidades o todos números de mensajes o agentes, como en las pruebas 17,19 y 24) fue debido al hecho que el método inteligente tenía propiedades deseables para solo ciertos valores de la calidad o números de agentes y esas propiedades se aplican para valores interesantes de la calidad y de números de agentes y de mensajes, tal como se explica en las tablas 8 y 10. Además, no podría ser siempre posible rechazar . Por ejemplo, aunque la hipótesis nula era que el modo inteligente alcanzaba una calidad mínima a una tasa mayor o igual que el modo simple, la fue rechazada a nivel de significación $\alpha \in \{0.01, 0.05\}$ para todas las calidades < 1.0 en la prueba 12, es decir, que el modo inteligente alcanza más a menudo la calidad = 1.0. Ésta es una propiedad deseable (alcanzar máxima calidad) lo que supera las otras ocasiones cuando el modo simple alcanza calidades más altas (ya que la suma de las frecuencias con que se alcanza una calidad debe ser 1, si el número de TOs es constante). La tabla 8 explica los resultados de esas pruebas.

Se ha visto que algunas propiedades eran mejores en el modelo inteligente cuando se tenían pocos agentes (menos de 30, ver prueba 7 en tabla 8 y prueba 19 en tabla 10) mientras que en otro caso lo era cuando los agentes eran más de 100 (ver sección 6.7.2). Esto se debe tal vez a los valores dados a los parámetros del sistema tal como se definen en la sección 5.4, que deberían ser ajustados en cada aplicación del modelo para obtener resultados deseables.

CAPÍTULO 7

FORMACIÓN DE HOLONES

En este capítulo se explica la aportación A2, es decir, la formación de estructuras holónicas a partir del funcionamiento de un sistema que implementa el modelo propuesto en esta Tesis.

7.1. Justificación teórica

En esta sección se muestra la correspondencia entre los conceptos manejados en los artículos mencionados en la sección 2.8.5.1 y esta Tesis. De esa forma, dado que existe una operación que permite agrupar agentes dentro de un holón en ambos casos (en el teorema de la sección 2.8.5.1.2 y aquí), queda probado que en el modelo presentado en esta Tesis puede generar holones. De forma adicional se mostrará que la dinámica del sistema lleva a la formación de holones y que con una cantidad de agentes suficientemente grande, la probabilidad de que haya estructuras no holónicas es despreciable. Se usará el símbolo “►” para abreviar “se corresponde por la construcción realizada con”.

- 1) $\mathcal{A}$ Conjunto de n agentes ► Conjunto de N *peers* $\{\alpha_1, \dots, \alpha_N\}$
- 2) ε : Conjunto de estados del entorno del sistema ► El estado del entorno del *peer* α_i , o sea ε_i , está definido por:
 - Indicador de mensajes limitados
 - Información de cada agente queriendo ingresar a través de α_i (CHECK.IN)

- Ω esperando respuesta
- los mensaje(s) de consulta enviado(s) al α_i por los agentes $\beta'_1, \beta'_2, \dots$
- las respuesta(s) enviada(s) al α_i por los agentes $\beta_1, \beta_2, \dots$
- los *time-outs* por mensajes que no le llegan a tiempo a α_i de parte de los agentes $\gamma_1, \gamma_2, \dots$
- los *time-outs* por mensajes que no le llegan a tiempo a $\gamma'_1, \gamma'_2, \dots$ de parte de α_i

El estado del entorno del sistema, ε_S , esta dado por (considerando que Ω no pertenece al sistema) por:

- Indicador de mensajes limitados
- Información de agentes queriendo ingresar al sistema (CHECK.IN)
- Mensaje(s) de consulta enviado(s) por Ω esperando respuesta
- Respuesta(s) enviadas a Ω

Por lo tanto es $\varepsilon \equiv \varepsilon_S$

- 3) Π función de percepción de los agentes $\Pi : \varepsilon \rightarrow P_1 \times \dots \times P_n$ ► Es lo que los distintos agentes perciben del entorno del sistema
- 4) S_i : estado del agente α_i ► Estado S_i definido por
 - Número de mensajes disponibles
 - Cantidad de *time outs* por agente consultado por α_i
 - Modo inteligente o simple de α_i
 - Consultas de otros pendientes de respuesta y origenes de ellas (identificador de sus originadores)
 - Consultas propias pendientes de respuesta y destinatarios de ellas
- 5) Φ_i : función del agente α_i $\Phi_i : S_i \times P_i \rightarrow S_i \times A_i$ ► La función (como actividad o funcionamiento) del agente α_i hace que pase de tener cierta cantidad de mensajes disponibles, asi como una cantidad de consultas a responder por él o que le deben responder y un cierto modo de trabajo a otra cantidad de mensajes, consultas etc. luego de realizar alguna acción
- 6) P_i : percepciones del agente α_i ► Las percepciones posibles para un agente son:
 - Mensajes que llegan (consultas y respuestas) y sus originadores
 - TIME OUTS producidos y sus originadores
 - Solicitudes de CHECK IN (incluyendo identificación de los originadores)
 - Indicación de mensajes limitados
- 7) A_i : conjunto de acciones del agente α_i ► Las acciones posibles de un agente en el sistema (en cuanto a lo que afecta su interacción con los demás) son:

- Responder consulta
 - Ignorar consulta
 - Hacer consulta
 - Responder solicitud de CHECK IN
 - Ignorar solicitud de CHECK IN
 - Pasar a modo inteligente
- 8) $\bar{\Delta}$: Cambio de estado del sistema $\bar{\Delta} : \varepsilon \times S_1 \times \dots \times S_n \rightarrow \varepsilon \times S_1 \times \dots \times S_n$ ► El sistema cambia de estado cada vez que o bien cambia el estado de su entorno o bien el de alguno de sus agentes
- 9) Π' : función de percepción con un holón formado. Construcción de Π' : Se construye tomando como percepción del holón un vector de percepciones como se vio en la Sección 2.2. ►
- Sea el holón α^* que se forma a partir de los agentes $\{\alpha_1, \dots, \alpha_k\}$:
 - Cuando un agente, para un campo dado, determina sus k mejores *peers*, se esta formando un holón (α^*) con cabeza ese agente y cuerpo esos k *peers*.
 - Los agentes pueden participar en varios holones: un agente puede ser de los mejores en un campo X para uno que consulta y puede ser de los mejores para el campo Y para otro
 - Se supone que Ω también elige los mejores agentes para su consulta
- La formación de un holón a partir de los agentes existentes para el modelo presentado en esta Tesis se puede plantear formalmente como en el cuadro adjunto más adelante, Cuadro 1.
- 10) $\bar{\Delta}'$: Cambio de estado del sistema con un holón definido ► Se construye igual que como se hizo en la sección 2.8.5.1.
- 11) Recurso abstracto: Cualquier dispositivo (incluyendo información) o herramienta que mejora el comportamiento de un agente ► Los recursos manejados por los agentes son:
- Energía (batería) – no modelada en sí
 - Mensajes disponibles
 - Predicciones de los otros agentes y sus calidades
 - Indicador de mensajes limitados
- 12) u : Función de utilidad. $u : E \times S \rightarrow \mathbb{R}$. Para el agente i -ésimo es $u_i : E_i \times S_i \rightarrow \mathbb{R}$ ► $Q(\alpha, m)$. La utilidad obtenida por un agente α_i se mide como Q : la calidad de la información que es capaz de devolverle al sistema (al que lo consultó)
- 13) E_i : entorno del agente α_i ► E_i comprende a los demás agentes y a Ω
- El conjunto de sus estados es ε_i definido previamente

14) Mecanismo de asignación de recursos:

- Es controlado por los miembros del holón a través de negociación individual y se corresponde por construcción con la asignación en el modelo: la asignación de recursos se da cuando el agente elige a los mejores colaboradores para un cierto campo. Técnicamente es un mecanismo basado en el mercado donde cada agente publica una tarea o recurso y los demás hacen ofertas por el recurso o ejecución de la tarea. Lo que se hace en el modelo es un *sealed-bid-first- Price auction* (la oferta más alta gana, sin importar quién la hizo) (Krishna, 2003).

Por último, la hipótesis mencionada en 2.8.1.5.2 (que todos los agentes se puedan comunicar entre sí y colaborar con todos los demás) se cumple aquí por ser un sistema P2P plano.

Dado el conjunto de agentes $\{\alpha_1, \dots, \alpha_k, \alpha_{k+1}, \dots, \alpha_N\}$

- La agrupación inicial H de $\{\alpha_1, \dots, \alpha_k\}$ se produce cuando se determinan por parte de Ω los k mejores agentes para su consulta. La cabeza de H será Ω y los $\{\alpha_1, \dots, \alpha_k\}$ cabezas de sus sub-holones. Ese holón tiene autonomía y se comunica a través de su cabeza.
- Dado el par $\{H, \alpha\}$ se forma H' cuando algún agente β de algún $H_i \subseteq H$ determina que α es uno de los k mejores agentes (de las mejores fuentes) para algún tipo de consulta. La cabeza será β y el cuerpo serán α y los agentes que él consulte (preferidos) más otros sub-holones que ya tuviera β . La comunicación se realiza a través de β con una probabilidad de que esto no se dé que decrece con N , ver fig. 23.
- Dados $\{H_1, H_2\}$ se forma H' cuando algún agente β que no pertenece ni a H_1 ni a H_2 determina que las cabezas de H_1 y H_2 están entre los k mejores fuentes para algún tipo de consulta. La cabeza de H' será β y H_1 y H_2 serán parte del cuerpo.

Cuadro 1. Formación de los holones

Dado el funcionamiento en fases explicado en 5.1 y considerando cómo se van formando estructuras (casi seguramente holónicas, como se discute a continuación) según el punto 9 anterior queda explícito que el sistema genera holones al funcionar. Si bien podría darse que dichas estructuras no se comunicasen únicamente a través de sus cabezas (ver figura 26) estas situaciones son extremadamente improbables al crecer el número de agentes

(para simplificar la notación lo llamaremos N), por eso se dice que las estructuras son casi seguramente holones. Esto se ve considerando las probabilidades individuales de que se produzca una situación como la de la figura 26. Suponiendo que los agentes son independientes en las consultas que realizan (o sea que lo que consulta uno no influye a lo que consulta otro) y que un agente no puede consultar al que le plantea la consulta, se tiene:

$P(\text{se dé la situación anormal de figura 26}) = P(\alpha \text{ tenga a } \gamma \text{ como preferido}) P(\alpha \text{ tenga a } \beta \text{ como preferido}) P(\beta \text{ tenga a } \gamma \text{ como preferido}) \approx 1/(N-1)^{3CK-1}$ ya que es:

$$P(\text{se elija } \gamma \text{ como preferido de } \alpha) = 1/(N-1)^C \quad (Ec. 14)$$

$$P(K \text{ agentes dados sean los preferidos de } \alpha) = 1/(N-1)^{C.K} \quad (Ec. 15)$$

$$P(\text{ocurra lo anterior para 3 agentes}) = \left(\frac{1}{N-3} \frac{1}{N-2} \frac{1}{N-1}\right)^{CK} \quad (Ec. 16)$$

$$P(\text{ocurra lo anterior en todo el sistema}) = P' = A_3^N \left(\frac{1}{N-3} \frac{1}{N-2} \frac{1}{N-1}\right)^{CK} = \frac{N(N-1)(N-2)}{((N-3)(N-2)(N-1))^{CK}} \leq \frac{N(N-1)(N-2)}{(N-3)^{3CK}} \quad (Ec. 17)$$

donde

- K es un parámetro del sistema = número de agentes a considerar como preferidos para un agente
- C es otro parámetro del sistema: el número de iteraciones que se hacen antes de elegir a un agente como favorito (llamado MENS.PREVIOS en la sección 5.4)
- $P(\bullet)$ denota la probabilidad, como es usual.

Se comprende que ya para valores relativamente bajos de C , K y N la probabilidad P' es muy pequeña. Además esa probabilidad es calculada suponiendo que los agentes disponen de los recursos necesarios como para realizar todas las comunicaciones necesarias, por lo cual la probabilidad final P' buscada es incluso menor:

$$P' \leq N(N-1)(N-2) \left(\frac{1}{N-3} \frac{1}{N-2} \frac{1}{N-1}\right)^{CK} \leq \frac{N(N-1)(N-2)}{(N-3)^{3CK}} \approx \frac{1}{N^{3CK-3}} \quad (Ec. 18)$$

Por ejemplo, para $N = 20$ agentes, $C = 3$ iteraciones y $K = 5$ preferidos es $P' \leq P \approx 1/20^{3 \times 3 \times 5 - 3} = 0.5E-42$, lo cual es despreciable.

Figura 26: Caso de comunicación anormal *Las flechas oscuras indican las comunicaciones con los preferidos, los hexágonos los holones. Entre γ y β hay una comunicación directa que debería ser a través de α .*

7.2. Caracterización de la auto-organización en el sistema

La formación de los holones es posible gracias a las propiedades de auto-organización del sistema. Se detallan a continuación cómo las distintas características de auto-organización señaladas en 2.10 se instancian en el modelo presentado. A continuación está indicada la correspondencia de cada punto.

Características básicas:

- **Límites:** las fronteras del sistema son determinadas por él mismo, la decisión sobre la afiliación de sus nuevos componentes debe ser hecha por el propio sistema.
 - En la arquitectura propuesta un agente no reconoce el CHECK-IN de otro nuevo si el número de mensajes del primero está por debajo de un cierto umbral. En versiones futuras se podría pedir además que el nuevo agente presentara algún certificado que comprobara su identidad o autenticidad.
- **Reproducción:** el sistema puede y debe reproducir su estructura (mediante la adición, remoción o cambio de un *peer*, sus datos o relaciones o conexiones con otros *peers*).

- Solo se cambian conexiones lógicas, es decir, con quien se comunica un agente, ya que no tiene sentido la replicación de datos.
- **Mutabilidad:** debe ser capaz de cambiar su estructura, por ejemplo cambiando las conexiones o formando *clusters*.
 - Es el cambio de las conexiones mencionadas, o pasar de modo irrestricto a inteligente
- **Organización:** el sistema se organiza ya sea como jerarquía, heterarquía o ambas.
 - En el presente caso es finalmente ambas (holones).
- **Métricas:** es capaz de detectar perturbaciones disparadas desde el entorno,
- **Adaptabilidad:** es capaz de reaccionar a las perturbaciones apropiadamente (re-estructurando agentes, incorporando redundancia, etc.).
 - La adaptación se hace a través de la nueva selección de mejores colaboradores.

7.3. Caracterización para la autonomía

Se describe a continuación cómo se instancian en el sistema las distintas características propias de la autonomía, que es otra característica básica para que se puedan formar los holones.

- **Feedback:**
 - La calidad de las respuestas dadas por los componentes y la decisión de responder o no son ejemplos de ellos en este caso
- **Reducción de la complejidad:**
 - Se forman aquí holones. Los holones permiten la reducción de la complejidad de un sistema debido al encapsulamiento que realizan de la información y el control.

■ **Aleatoriedad:**

- Se utiliza aleatoriedad en distintas partes: en el caso en el que el sistema hace una elección de tipo el temple simulado (cuando decide si contestar a pesar de tener una respuesta de menor calidad que el que consulta) o cuando se eligen los campos a consultar y si se consulta o no por ese campo, etc.

■ **Auto-organizado críticamente:**

- La falla en alguna de las conexiones lleva a toda una serie de nuevos caminos de consulta. Cuando un *peer* no responde (lo que se puede considerar una falla de la conexión) se selecciona el siguiente *peer* de la lista de preferidos, que a su vez tendrá otros preferidos, por lo cual la consulta viajará por nuevos caminos

■ **Emergencia:**

- Por ejemplo, si se tienen en cuenta las fórmulas (7) a (10) se ve que el sistema puede dar un resultado con mejor calidad que cualquiera de los componentes por separado, tomando un $\hat{N}$ lo suficientemente grande y una cantidad de mensajes grande (MAX.NRO.MENSAJES.AG) de forma que para un mensaje m dado y α arbitrario (ver sección 5.2) es $Q(\alpha, m) > \overline{Q}(\alpha, m)$.

■ **Otras características:**

- La igualdad de los *peers* contrarresta el riesgo de los únicos puntos únicos de falla. Todos los *peers* son iguales aquí, salvo en entrenamiento que hayan tenido (Fase 0, ver sección 5.3).

■ **Un sistema P2P debería ser idealmente capaz de reproducir su estructura, no solo sus datos. Esa reproducción debería ocurrir automáticamente y no disparada manualmente o provenir desde afuera.**

- El sistema detecta automáticamente demoras en responder de los agentes y los agentes empiezan a trabajar en modo inteligente (holón).

7.4. Ejemplo de formación de holones

Se describe a continuación un ejemplo que busca clarificar el funcionamiento de una implementación del modelo y mostrar cómo se da la formación de holones. Los datos usados en este ejemplo están inspirados en el juego de datos para el caso B (clasificar si un automóvil es aceptable o no basándose en un conjunto de atributos del mismo), pero con una modificación en su estructura, como se verá.

7.4.1. Dinámica del sistema

Sea Ω el agente originador que envía un mensaje. Se toma como conjunto de agentes de este ejemplo a: $\{\alpha, \beta, \gamma\}$. A los efectos de hacer manejable la cantidad de pasos a dar el número máximo de mensajes permitido por agente es 10.

Los registros que se manejarán tienen campos A, B, C, D tales que

- A se forma a partir de B y C.
- C se forma a partir de D.
- D a partir de E, F y G.
- B depende de G.

Es decir $A=f(B,C)$, $C=g(D)$, $D=h(E,F,G)$ y $B=p(g)$ siendo $f, g \dots p$ ciertas funciones.

Sea M el mensaje completo que envía Ω o sea $M=A||B||C||D$ donde $||$ representa la concatenación como es usual. Más concretamente, los distintos M que llegan son tuplas de datos relacionadas con el confort que tienen vehículos de distintos tipos y precios. En ese sentido se toma:

- A= valoración del confort (entero de 1 al 5, 5 es el máximo).
- B= nro. de puertas (2 a 5).
- C= existencia de aire acondicionado (S/N).
- D= precio de venta (en dólares norteamericanos).
- E= tipo de vehículo (utilitario: si o no).

- F= origen del vehículo (nacional o importado. 1 = nacional).
- G= modelo (sedán, *coupé*, etc.).

En el juego de datos usado finalmente se convirtieron todos los campos a numéricos.

La incertidumbre que puede tener un agente respecto a un campo existe solo para los campos compuestos (es decir, que dependen de otro u otros; para los campos que no son compuestos o dependientes de otros no se considera ningún error.) y es el error “de predicción” del agente para ese campo. A su vez, el error de predicción para el mensaje M es el de clasificar erróneamente el registro decidir si un vehículo brinda el suficiente confort para las características que tiene (decisión binaria).

La calidad de información (como porcentaje de error de predicción, en el caso de suponer los datos sin vaguedad) para cada campo compuesto y cada agente obtenidos luego de una etapa de entrenamiento se toman como lo indica la tabla 12.

	A	B	C	D	M
α	5%	10%	15%	20%	5%
β	25%	10%	10%	15%	10%
γ	10%	15%	20%	15%	25%

Tabla 12: Calidad de la información de cada agente para cada campo

La tabla 12 significa que, por ejemplo, el error cometido por el agente α al predecir A es del 5 %, suponiendo los datos *crisp*, sin vaguedad.

Como se describió en la sección 5.3 “Dinámica del sistema”, existe un periodo (tiempo de estabilización) durante el cual Ω envía un mensaje a todos los agentes y a partir de cierto momento (que él decide) solo le envía a los que han respondido más veces con el menor error. Lo mismo sucede con todos los otros agentes.

Se supone que cada agente decide mandarle el mensaje al agente que le ha respondido mejor más veces esperando para ello haber recibido sólo 1 respuesta (o sea, no manda 20 mensajes y luego ven las frecuencias de éxito de cada agente y elije el de mayor frecuencia, sólo manda 1 mensaje y ve cual es el agente más exitoso). También para simplificar se tomó la calidad de un campo inversamente proporcional al error de predicción de ese campo sin considerar las demás variables (número de reenvíos de la consulta, número de agentes consultados, etc.).

Cuando un agente desea consultar a otro sobre un campo X, consulta a todos los que tienen un número de éxitos iguales al máximo cuando él los ha consultado previamente para ese campo (es decir, consulta a los que le han respondido con mayor calidad). Se entiende que cada agente (incluyendo Ω) para ello lleva una tabla con esas frecuencias (Ω lleva sólo de M): para Ω la tabla “MEJOR” recoge el número de veces que cada agente dio la mejor respuesta (ver tabla 13), para los demás agentes esa información se lleva en la tabla 14 “MEJORES” .

Campo	α	β	γ
M	0	0	0

Tabla 13: Tabla MEJOR

MEJOR(β)=3 significaría que hasta el momento β fue el agente que dio tres veces la mejor respuesta al mensaje de Ω .

Téngase en cuenta que M más que un campo es todo el mensaje, $M=A||B||C||D||E||F||G$.
 $M= \text{confort}||\text{nro. puertas}||\text{aire acondicionado}||\text{precio}||\dots$

Para α	α	β	γ	Para β	α	β	γ	Para γ	α	β	γ
A				A				A			
B				B				B			
C				C				C			
D				D				D			
M				M				M			

Tabla 14: Tabla MEJORES

Originalmente, antes de que recibir ningún mensaje todos estos campos de la tabla están en 0.

También se lleva el número de mensajes restantes por agente en la tabla 15 (la tabla va variando con el tiempo):

Agente	Número de mensajes
α	10
β	10
γ	10

Tabla 15: Tabla MENSAJES-RESTANTES

(o sea, para α hay 10 mensajes restantes en ese momento).

Cada agente puede consultar o no a otro agente sobre cierto campo, con una probabilidad proporcional a su calidad (error de predicción) para ese campo (o sea, a gran error, o pequeña calidad, gran duda y gran probabilidad de que consulte). Nótese que el error de predicción es el determinado en tiempo de entrenamiento, FASE 0).

La fusión de las distintas respuestas que puede recibir un agente consiste, en este ejemplo, en elegir la de menor error.

Los cuadros mencionados antes evolucionan con el tiempo: Para Ω , el cuadro de número de veces que cada agente dio la mejor respuesta al “campo” M evoluciona según la tabla 16:

t	α	β	γ	t	α	β	γ
1	0	0	0	26	3	0	0
2	0	0	0	27	3	0	0
3	0	0	0	28	3	0	0
4	0	0	0	29	3	0	0
5	0	0	0	30	4	0	0
6	0	0	0	31	4	0	0
7	0	0	0	32	5	0	0
8	0	0	0	33	5	0	0
9	0	0	0	34	6	0	0
10	0	0	0	35	6	0	0
11	0	0	0	36	6	0	0
12	0	0	0	37	6	0	0
13	0	0	0	38	6	0	0
14	1	0	0	39	6	0	0
15	1	0	0	40	6	0	0
16	1	0	0	41	6	0	0
17	1	0	0	42	6	0	0
18	1	0	0	43	6	0	0
19	1	0	0	44	6	0	0
20	2	0	0	45	6	0	0
21	2	0	0	46	6	1	0
22	2	0	0	47	6	1	0
23	3	0	0	48	6	1	0
24	3	0	0	49	6	1	0
25	3	0	0	50	6	1	1

Tabla 16: Evolución de la tabla “MEJOR”

significando que, por ejemplo en el instante $t = 25$, el agente α era el que había respondido 3 veces a la consulta de Ω mientras que los demás no habían respondido nunca aún, por lo que α es el mejor agente (en el instante 25) según Ω . Para los otros agentes la evolución es como se muestra en la tabla 17.

Para α	α	β	γ	Para β	α	β	γ	Para γ	α	β	γ
A	1 (en t=6)	1 (en t=18)		A	1 (en t=9)			A			
B				B				B			
C				C				C			
D				D				D			
M				M	1 (en t=10)			M			

Tabla 17: Evolución de la tabla “MEJORES”

Produciéndose los cambios en los instantes señalados como (en $t = xx$); es decir, que en $t = 18$ para α el agente que mejor respondió su consulta sobre el campo C fue β . El número de mensajes restantes por agente evoluciona con el tiempo según la tabla 18.

t	α	β	γ	t	α	β	γ
1	10	10	10	26	2	4	10
2	10	10	10	27	2	3	10
3	10	10	10	28	2	2	10
4	10	10	10	29	2	2	10
5	9	10	10	30	2	2	10
6	8	8	10	31	2	2	10
7	8	6	10	32	1	2	10
8	8	6	10	33	1	2	10
9	8	5	10	34	0	2	10
10	8	5	10	35	0	2	10
11	8	5	10	36	0	2	10
12	8	5	10	37	0	2	10
13	7	5	10	38	0	2	10
14	7	5	10	39	0	0	10
15	7	5	10	40	0	0	9
16	6	5	10	41	0	0	9
17	5	5	10	42	0	0	8
18	5	4	10	43	0	0	8
19	5	4	10	44	0	0	8
20	4	4	10	45	0	0	7
21	4	4	10	46	0	0	7
22	4	4	10	47	0	0	7
23	3	4	10	48	0	0	7
24	3	4	10	49	0	0	7
25	3	4	10	50	0	0	7

Tabla 18: Evolución de “MENSAJES-RESTANTES”

O sea, que por ejemplo en el instante 25 el agente α tiene 3 mensajes restantes mientras que β tiene 4.

El número de mensajes restantes para cada agente no es conocido por los demás pero sí conocen el propio; los agentes demoran un tiempo aleatorio en responder.

La figura 27 muestra un diagrama inspirado en el AUML (*Agent Unified Modeling Language*, (Padgham & Winikoff, 2005)) que refleja gráficamente la evolución del sistema, en el eje vertical está representado el paso del tiempo mientras que las distintas bandas verticales representan a los distintos agentes, las flechas continuas mensajes de consulta y las cortadas respuestas a consultas. Los hitos más importantes de la figura 27 son

presentados en la figura 28 donde el tiempo transcurre de arriba hacia abajo.

En la figura 27 es:

- Las flechas discontinuas representan respuestas, las continuas, solicitudes (consultas). Los trazos verticales significan que un mensaje generó otros como consulta.
- $P(Y,Z)$ significa “predicción, hecha por Y, del campo Z”.
- Las letras en los casilleros, el campo que se consulta.
- El tiempo transcurre verticalmente, de arriba abajo. Cada fila representa un instante de tiempo (t), numerado 1,2... .
- Una $\emptyset$ significa no hay respuesta (por tener un error mayor que el que consulta), y
 - si no responde (por no tener mensajes).
- Los M_i representan mensajes ($M_i \neq M_j$ con $i \neq j$).
- Un holón H con un agente U como cabeza se forma cuando quien consulta a H tiene a U como primera opción al mandar el mensaje de consulta. A partir del momento en que U se convierte en la mejor opción, decimos que emergió H.
- Un agente responde la consulta solo si su error de predicción es estrictamente menor que el del consultante, y si tiene mensajes.
- Al comenzar a operar, si Ω aún no tiene al menos dos respuestas para decidir cuál es mejor, espera a tenerlas o a que se produzca un *time-out*.

La figura 29 muestra cómo va evolucionando el sistema en este ejemplo.

Figura 27: Diagrama de la dinámica del ejemplo

<i>T</i>	<i>Comentario</i>
1	α consulta por el campo Confort a β y γ . $M(\alpha)=10$, $m(\beta)=10$, $m(\gamma)=10$
2	β consulta por Nro de puertas y Precio de venta
4	Los agentes demoran un tiempo aleatorio en responder, por eso aquí no sucede nada.
5	β tiene peor error para Confort que α (10% vs. 5%) y entonces no contesta. $m(\alpha)=9$
6	Lo mismo sucede con γ (10% vs. 5%). $MEJORES(\alpha, Confort, \alpha) + 1$ $m(\alpha)=8$
8	α y γ no tienen mejores errores para nro. puertas y Precio que β y no contestan (10% de β vs. 10% y 15 % ; 15% de β vs. 20% y 15% de α y γ respectivamente). $m(\beta)=8$
9	$m(\beta)=6$. $MEJORES(\beta, nro\ puertas, \beta) + 1$. $MEJORES(\beta, precio, \beta) + 1$
11	β da su predicción para M junto con su indicación de calidad (error): S (Adecuado) . Error=10% . $m(\beta)=5$
13	Idem α : responde N (no adecuado). Error=5% . $m(\alpha)=7$
14	Ω decidirá qué clasificación hace del vehículo a partir de las respuestas. $MEJOR(\alpha)+1$ (y queda en 1)
15	Las consultas subsiguientes se hacen a α por ser el de mejor desempeño, mientras conteste.
16	α consulta por el campo Aire acondicionado . $m(\alpha)=6$
17	y no contesta ya que el error de γ para Aire acondicionado = 20% y error α para ese campo es 15%. $m(\alpha)=5$
18	β da su predicción para Aire acondicionado junto con su indicación de calidad (error). Error= 10%. Aire acondicionado = S . $MEJORES(\alpha, aire\ acondicionado, \beta) + 1$
20	α da su predicción para m: Confort = N (no adecuado). Error = 5% . $MEJOR(\alpha)+1$ (y queda en 2). $m(\alpha)=4$
23	α da su predicción para m: Confort = S (adecuado). Error = 5% . $MEJOR(\alpha)+1$ (y queda en 3). $m(\alpha)=3$
25	α consulta a δ que a su vez consulta a γ por aire acondicionado
26	$m(\alpha)=2$
27	$m(\delta)=3$
28	El error de γ es mayor que el de δ (20% vs 10% respectivam.) por lo que γ no contesta. $m(\delta)=2$
29	δ da su predicción de aire acondicionado = S ya que error de α en aire acondicionado = 15%.. Error de δ = 10%
30	Predicción de α para M4: Confort = N (no adecuado). Error = 5% . $MEJOR(\alpha) + 1$ (queda en 4)
32	Predicción de α para M5: Confort = S (adecuado). Error = 5% . $MEJOR(\alpha) + 1$ (queda en 5)
34	Predicción de α para M6: Confort = N (no adecuado). Error = 5% . $MEJOR(\alpha) + 1$ (queda en 6). $m(\alpha)=0$
35	A partir de aquí α s/mensajes. Se le preg a los que le siguen en frecuencia (β y γ)
37	Le manda a otros con frecuencia de respuesta menor que α (β y γ)
*39	δ consulta a γ y a α por Confort . $m(\delta)=0$
40	Error de γ para Confort = 10%. Predicción de γ para Confort = 3 . $m(\gamma)= 9$
42	El agente γ consulta a α por Nro. De puertas . α no contesta porque no tiene más mensajes. $m(\gamma)=8$
44	β responde y queda sin mensajes. Predicción de δ para M6: Confort = S (adecuado). $m(\delta)=4$. Error = 10%
45	Este mensaje es desestimado por α por estar fuera de tiempo. Los agentes demoran un tiempo aleatorio en responder. Predicción de γ para M6: Confort = S . Error = 25% . $m(\gamma)=7$
46	$MEJOR(\delta) + 1$ (y queda en 1)
49	Se le pregunta al siguiente de la lista con más respuestas exitosas (γ)
50	Predicción de γ para M6: Confort = S . Error = 25% . $MEJOR(\gamma) + 1$ (y queda en 1). $m(\gamma)=6$

Figura 28: Cronología de los eventos del ejemplo

7.4.2. Formación de los holones

En el ejemplo desarrollado, el esquema de la figura 29 muestra las etapas de formación y disolución de los holones. En la figura 30 se muestra otra posible evolución de los holones, por ejemplo por usarse parámetros con distintos valores.

Figura 29: Evolución de los holones

Figura 30: Otra posible evolución de los holones. Los agentes que reciben la consulta de OMEGA son α, β, γ . En la parte superior: se tomó $MENS.PREVIO=1$. Desde el comienzo conocen el límite de mensajes por agente. No se producen ingresos en el periodo estudiado. $N=1$. En la parte inferior: Idem, pero $N=2$. La emergencia se produce en otros instantes.

Los instantes indicados 14, 36, 46, 49 señalan el momento en el que comienza cada fase y en el que el agente (u holón) Ω empieza a dirigir sus mensajes a un agente distinto del que venía haciéndolo. Por ejemplo, en el instante 14 pasa a dirigirse solo a α , y en el 46 solo a β . En el 36 α ya no tiene mensajes y no responde, por lo cual Ω se dirige a los otros agentes hasta poder seleccionar uno de nuevo. Ver la evolución en el tiempo de la tabla 13 - MEJOR.

Cada iteración (cada nueva consulta que hace Ω) puede presentar las dos fases, FASE 1 y FASE 2 bien definidas. En la iteración 1 en la fase FASE 1 se observa al conjunto de todos los agentes trabajando en forma P2P. Posteriormente, luego de transcurrido el tiempo de estabilización, Ω ya sabe a quién dirigirse, se está en la FASE 2 (en caso de que Ω ya supiera a quien dirigirse, sólo sería la FASE 2), ya tiene el agente que mejor le ha respondido y es entonces cuando se puede decir que se ha generado un holón. En este caso la emergencia del holón (ver figura 24) se produce en el instante 14 y luego

desaparece en el 36. En los instantes 14 y 46 es donde se evidencia el comportamiento inteligente del sistema.

En la situación 1 se puede apreciar que el agente α actúa como cabeza de un holón, siendo la puerta de entrada/salida de las consultas que se hacen al grupo y coordinando los resultados de los otros dos (β y γ). A su vez estos β y γ actúan formando el cuerpo, los que procesan otras partes del mensaje. Lo mismo se puede decir en la situación 3, con α como cabeza y γ como cuerpo.

Nótese como se alternan las estructuras P2P con holónicas al ir reduciéndose los mensajes. Son holones porque son conjuntos de agentes semi-autónomos, tal vez heterogéneos, agrupados por niveles de forma que se puede apreciar una estructura cabeza-cuerpo en cada nivel, para alcanzar un fin común. Utilizan información propia de su nivel para realizar sus tareas.

Asimismo se cumplen las condiciones que debe cumplir una holarquía según (Gerber *et al.*, 1999):

- 1) **Autonomía:** los agentes que responden consultas son autónomos y pueden o no responder. Al participar en un holón pierden cierta autonomía ya que deben realizar tareas para el mismo durante cierto tiempo.
- 2) **Comportamiento orientado a la meta común:** los agentes cooperan para responder con una información de la mayor calidad posible a una consulta que llega.
- 3) **Capacidades de grupo ampliada:** El holón responde con una calidad que incorpora información que no pueden tener sus integrantes por separado (por ejemplo, número de agentes que contestaron, variación de las calidades, etc.). La calidad puede ser menor que la de los consultados, debido al temple simulado que se realiza (ver sección 5.5).
- 4) **Creencias:** los agentes tienen representaciones explícitas del entorno (tablas de preferencia, lista de agentes activos, etc.).
- 5) **Racionalidad limitada:** el súper-holón monitorea el uso de recursos por parte de los sub-holones. En este caso no se controla en forma directa el uso de recursos aunque se puede decir que lo hace indirectamente: si un agente A tiene menos de

x % de mensajes, sortea si consultar o no. Al no consultar, se evita que el sub-holón H1 de A gaste recursos en responder ya que puede darse que H1 responda y el consultante (A) finalmente no pueda responder a su vez por falta de mensajes.

- 6) **Comunicación:** las comunicaciones entre sub-holones pasan por la cabeza necesariamente. Cuando se han conformado los holones, las comunicaciones solo pasan por las cabezas. La probabilidad de que un holón deje de comportarse como tal y se comunique por fuera de su cabeza suele ser ínfima, como se vio en la sección 7.1.

CAPÍTULO 8

CONCLUSIONES

En esta tesis se ha propuesto un modelo de red P2P, el e-P2P, basado en una red P2P plana y no estructurada (“pura”) formada por agentes colaborativos que se mueven dentro de un Sistema Inteligente Global o GIS la cual en algún punto comienza a evolucionar (de ahí el nombre de e-P2P o *evolvable* P2P). Que la red sea plana y desestructurada significa que no hay un mecanismo centralizado de forma de coordinar las operaciones del sistema, tal como un servidor central o “super-nodo”, que sirve a los otros nodos; en ella (y en el modelo e-P2P) todos los participantes tienen los mismos derechos y obligaciones y cualquier *peer* puede abandonar la red sin tener ello un impacto significativo en la operación del sistema. Los agentes participantes colaboran para atender un requerimiento (por ejemplo, una consulta) de un agente externo al GIS, denominado Ω , que en cierto momento empieza a interactuar con ese espacio. La organización de sus elementos se realiza tomando en cuenta la calidad de la información circulante. Una gran ventaja de las redes P2P es que se adaptan rápidamente a poblaciones que crecen velozmente ya que la consulta es enviada a todos los agentes de la red. La desventaja es que la consulta es planteada a numerosos *peers* sin preocuparse de si pueden o no responder. Esto causa congestión de la red y un uso exponencial de mensajes. Sin embargo, el enfoque propuesto aquí evita la congestión ya que consulta sólo los mejores *peers* que el agente haya encontrado para un campo en vez de todos los agentes. y el parámetro MENS.PREVIOS deben ser elegidos apropiadamente para este propósito. La escalabilidad es mejorada:

aunque todos los nodos sean consultados, no se encuentran tantos resultados potenciales como en redes no estructuradas y puras (por ej. Gnutella); el algoritmo de temple simulado evita quedar atrapado en óptimos locales facilitando la serendipia.

El modelo tiene dos diferencias claves respecto a una red P2P pura y sin ninguna inteligencia. Primero, en el caso del P2P pura, la consulta se hace sistemáticamente a todos los agentes, sin tener agentes preferidos, mientras que en el e-P2P la consulta se realiza a (casi) todos los agentes al principio y luego solo a los agentes preferidos. Segundo, mientras que en la P2P pura no se toma en cuenta si la calidad es mejor que la del que consulta a la hora de responder, en el e-P2P se responde, salvo en casos excepcionales, sólo si la calidad es mayor que la del que consulta. Por otra parte, el modelo e-P2P puede comportarse como una P2P pura si la cantidad de mensajes por agente es suficientemente grande. Estas características evitan un punto de fallo único ya que ningún *peer* es indispensable para el sistema. La red es, por lo tanto, altamente inmune a la censura, fallas técnicas y ataques maliciosos (Meer & Koppen, 2005a). El modelo propuesto conserva e incluso mejora estas propiedades con respecto, por ejemplo, al consumo del ancho de banda (medido como el número de mensajes usados) y el tiempo de respuesta (cuan a menudo es menor que el límite, es decir, no es un *time-out*). Sin embargo, el modelo e-P2P propuesto tiene un costo asociado: se tienen que mantener la información de direccionamiento (tablas de *routing*) para cada agente, tales como cuáles son los mejores agentes hasta el momento para cada campo y para cada agente y el proceso adicional que esto implica. Esto puede significar un problema cuando los *peers* entran al o dejan el sistema extremadamente rápido (a tasa muy alta) ya que tales tablas deberían ser actualizadas permanentemente. Este problema no es tomado en cuenta por considerar la hipótesis de que no hay *churn*.

Las diferencias mencionadas anteriormente hacen que el sistema implementado tenga un comportamiento especial en cuanto a su rendimiento. Al respecto, de los resultados cuantitativos obtenidos para los casos de estudio, se concluye que la RQ1 puede ser respondida afirmativamente estando esto soportado estadísticamente por al menos en el 80 % de todas las pruebas (ese porcentaje se calcula a partir de la cantidad de veces que las hipótesis correspondientes pudieron ser validadas). Ese porcentaje tal vez podría ser aumentado ajustando adecuadamente los parámetros del sistema.

Por lo tanto, dado que la respuesta a la RQ1 es afirmativa se puede decir que **se validó la hipótesis H1:** *hay una mejora significativa en el rendimiento del sistema medida como tráfico medio (mensajes recibidos y enviados) por agente para distintas calidades de respuesta y como la tasa de éxitos estrictos, mientras que el número de veces que el sistema no responde a tiempo es menor en el modo inteligente que en el simple.*

Por otra parte, se demostró que se pueden generar holones y cómo la probabilidad de que no se cumpla con la estructura de holón cabeza-cuerpo tiende a 0 muy rápidamente al crecer el número de agentes; por tanto, se puede contestar también afirmativamente a la RQ2. **De todo lo anterior se desprende que ambas RQ han sido respondidas afirmativamente.**

La respuesta a la RQ2 se dio mostrando la correspondencia entre los conceptos y entes manejados en el sistema y los del trabajo de Gerber *et al.* (Gerber *et al.*, 1999) de modo que se pudo demostrar que un sistema que implemente este modelo puede generar holones, y que la probabilidad de que haya estructuras que no sean holarquías tiende a 0 rápidamente al crecer el número de agentes, por lo que **se validó la hipótesis H2:** *el modelo posibilita la auto-organización de la red de forma de reaccionar con flexibilidad a los cambios en su entorno y en su composición, manteniendo su estabilidad y su calidad de predecible, más específicamente, que conduce a la formación de estructuras holónicas.* Esta formación era previsible ya que el modelo fue diseñado de manera de propulsar la auto-organización (y se muestran cómo se cumplen las condiciones para que la haya) y la generación de holones. Esos holones pueden ser vistos como entidades que generan informes los que se caracterizan por su mayor contenido semántico que los datos originales. En ese sentido cada par (*dato*, *calidad del dato*) puede ser considerado un informón de mayor nivel semántico (más rico semánticamente) que el *dato*.

Los holones poseen una serie de ventajas como la semi-autonomía, resistencia y recuperación ante fallas u otras alteraciones internas o externas, permiten distribuir la carga de control, etc. Si bien se sabe que se generan holones, conocer *a priori* cuáles van a ser resulta imposible debido a la aleatoriedad del comportamiento de los agentes. Los holones son pues el resultado de la auto-organización de los agentes.

Por último, si se tienen en cuenta la teoría del *sensemaking* en las organizaciones sociales de Dervin y Weick según la cual los individuos en una organización se estructuran

de forma de darle sentido a la información que reciben; entendiendo como una de las funciones asociadas a dar sentido la reducción de la vaguedad o incertidumbre que se encuentra en un mensaje para poder dar la respuesta más adecuada, y dado que la métrica de calidad usada penaliza dichas características, está claro que hay una correspondencia entre los objetivos del modelo propuesto y el *sensemaking*. Se utilizó fusión de información ya que es la manera natural de combatir la vaguedad y ambigüedad y es la que mejor se adapta a los métodos colaborativos de *sensemaking* y que además permite tener agentes heterogéneos realizando el procesamiento de la información de diferentes maneras.

Como líneas de trabajo futuro se consideran las siguientes:

- En el modelo propuesto un agente nunca sabe si su respuesta fue la mejor o si llegó fuera de hora (*time-out*). Agregar *feedback* sobre ello de parte del emisor (el que realiza la consulta) aumentaría los mensajes a mandar, sin embargo, se dispararía un comportamiento inteligente del receptor más tempranamente (el agente, al ver que está produciendo resultados de baja calidad, decide pasar a modo inteligente), ahorrando mensajes (por ejemplo al llevar a consultar el receptor a su vez a otros agentes para tener una mayor calidad en la respuesta). Analizar si se permite ese *feedback* con el consiguiente aprendizaje adicional de los agentes es entonces un tema de diseño a evaluar.
- Se ha supuesto que, dado un campo, el agente no tiene duda sobre cómo descomponerlo (o, equivalentemente, a partir de qué otros campos se obtiene). Es más, para poder razonar sobre situaciones, intenciones y amenazas se necesita una ontología compartida (Little & Rogova, 2009). Un tema a investigar es el caso de tener varias descomposiciones factibles entre las cuales se tiene que elegir una. En forma más general, se ha supuesto que el criterio de descomposición (vago o no) de los mensajes y campos es compartido por todos los agentes (por ejemplo, por usar una misma ontología). Se podría estudiar qué variaciones hacer al modelo cuando esto no sucede (por ejemplo, campos que no son conocidos por el agente consultado).
- Modificar el modelo de forma de cumplir con la condición de límite (*boundary*)

(ver sección 7.2) de forma que el sistema determina sus límites y el propio sistema toma la decisión de dejar ingresar o no a nuevos componentes. En el modelo propuesto un agente que tiene un número de mensajes (recursos, en su forma más general) por debajo de un umbral específico no reconoce la solicitud de ingreso (CHECK-IN) proveniente de un nuevo agente. En futuras versiones, el nuevo agente tal vez deba enviar algún tipo de certificado de forma de asegurar su identidad y autenticidad de la solicitud.

- Permitir un *time-out* distinto por agente (y así tener agentes “nerviosos” y otros “calmos”, en el sentido que los agentes con menor TO van a desechar las respuestas que les lleguen antes que los otros con mayor TO actuando como personas con prisa o nerviosas que no esperan hasta que llegue una respuesta).

- Realizar una evaluación de la usabilidad de una métrica de IQ definida como un real en $[0,1]$ para seres humanos puede ser de vital importancia, ya que de ello depende que el modelo sea adoptado en distintas aplicaciones, o, más específicamente, qué tipos de usuarios o en qué condiciones encuentran adecuado esa métrica.

- Sistemas de este tipo pueden ser aplicados en el manejo de grandes volúmenes, tales como en el caso de *Big Data*, de datos donde cada agente se especializa en el procesamiento de un subconjunto de esos datos. Dado que en el proceso de *big data* se distribuyen los datos entre distintos repositorios así como su procesamiento parece natural que en el modelo los distintos agentes se especialicen en procesar distintos subconjuntos de datos. El propio entrenamiento inicial (FASE 0) puede ser hecho de forma de especializar a los agentes (algunos agentes trabajan mejor con ciertos conjuntos o combinaciones de rangos de datos y otros con otros.), o, equivalentemente, se puede pensar que los agentes que ingresan al sistema ya están especializados en algún subconjunto o vista de los datos. Los agentes que entran y los que salen podrían también estar asociados a los nuevos datos que se van generando. Otro caso de agentes especializados por regiones de los datos (que podría coincidir con el caso de geolocalización) sería el de un sistema abierto de recomendación, donde los agentes recomiendan con más certeza lugares o actividades que les son cercanas geográficamente. Se podría pues estudiar cómo este modelo de sistema puede aportar a algunos de los problemas abiertos en el área de *big data*.

- Definir las propiedades que debe reunir una métrica de IQ para que el modelo mantenga sus bondades. El modelo presentado permite obtener sistemas con ciertas ca-

racterísticas considerando la IQ definida de cierta forma. Sería interesante estudiar qué propiedades debe tener una métrica de IQ para que un sistema de este tipo que la use en lugar de la aquí definida siga presentando las características encontradas.

- Sistemas con un modelo similar al presentado en esta Tesis se han analizado desde el punto de vista del control distribuido, en los que las consultas no se realizan salvo que la calidad de información esté debajo de un umbral constante (Wang & Hovakimyan, 2012). Se podría analizar cómo el modelo e-P2P puede ser usado en ese campo.

- A. Jalobeanu, J.A. Gutiérrez. 2006. Multisource data fusion for bandlimited signals: A Bayesian perspective. *25th workshop on Bayesian Inference and Maximum Entropy methods (MaxEnt'06)*.
- Abdelhamid, Neda, Ayesh, Aladdin, & Thabtah, Fadi. 2014. *Phishing detection based Associative Classification data mining*. Vol. 41.
- Abidi, M. A., & González, R. C. 1992. *Data Fusion in Robotics and Machine Intelligence*.
- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., & Cayirci, E. 2002. Wireless sensor networks: a survey. *Computer Networks*, **38**(4), 393–422.
- Alberts, David S., & Hayes, Richard E. 2005. *Power to the Edge. Command and Control in the Information Age*.
- Alex, Hitha, Kumar, Mohan, & Shirazi, Behrooz. 2008. MidFusion: An adaptive middleware for information fusion in sensor network applications. *Information Fusion*, **9**(3), 332–343.
- Alhoniemi, Esa, Hollm, Jaakko, Simula, Olli, & Vesanto, Juha. 1999. Process Monitoring and Modeling Using the Self-Organizing Map. *Integr. Comput.-Aided Eng.*, **6**(1), 3–14.
- Alonso, F., López, G., Pazos, J., Rodríguez, Patón, Silva, A., & Soriano, J. 2004. *Fundamental Elements of a Software Design and Construction Theory Informons and Holons*.
- Alston, Anthony, Beaument, Patrick, & Dodd, Lorraine. 2005a. Implementing Edge Organizations: Exploiting Complexity (Part 1: A Framework for the Characterization of Edge Organizations and their Environments).
- Alston, Anthony, Beaument, Patrick, & Dodd, Lorraine. 2005b. Implementing edge organizations:exploring complexity (Part I: a framework for the characterization of edge organization and their environments. *10th International Command and Control Research & Technology Symposium*.
- Amgoth, Tarachand, & Jana, Prasanta K. 2015. Energy-aware routing algorithm for wireless sensor networks. *Computers & Electrical Engineering*, **41**, 357–367.

- Amiri, Ali. 2006. Designing a distribution network in a supply chain system: Formulation and efficient solution procedure. *European Journal of Operational Research*, **171**(2), 567–576.
- Andrade, J., Ares, J., García, R., Martínez, M. A., Pazos, J., & Suárez, S. 2015. A Game Theory Based Approach for Building Holonic Virtual Enterprises. *IEEE Trans. on Syst., Man, and Cybern.: Syst.*, **45**(2), 291–302.
- Antonopoulos, Nick. 2010. *Handbook of Research on P2P and Grid Systems for Service-Oriented Computing: Models, Methodologies and Applications: Models, Methodologies and Applications*.
- Auslaender, O. M., Steinberg, H., Yacoby, A., Tserkovnyak, Y., Halperin, B. I., Baldwin, K. W., Pfeiffer, L. N., & West, K. W. 2005. Spin-Charge Separation and Localization in One Dimension. *Science*, **308**, 88–92.
- Ayyub, B. M., & Klir, G. J. 2006. *Uncertainty Modeling and Analysis in Engineering and the Sciences*.
- Azadeh, A., Saberi, M., & Asadzadeh, S. M. 2011. An adaptive network based fuzzy inference system-auto regression-analysis of variance algorithm for improvement of oil consumption estimation and policy making: The cases of Canada, United Kingdom, and South Korea. *Applied Mathematical Modelling*, **35**(2), 581–593.
- Azharuddin, Md, Kuila, Pratyay, & Jana, Prasanta K. 2015. Energy efficient fault tolerant clustering and routing algorithms for wireless sensor networks. *Comput. Electr. Eng.*, **41**(C), 177–190.
- Aziz, Ashraf. 2011. A Soft-Decision Fusion Approach for Multiple-Sensor Distributed Binary Detection Systems. *IEEE Trans. Aerospace and Electronic Systems*, **47**, 2208–2216.
- Aziz, Ashraf M. 2008. A simple and efficient suboptimal multilevel quantization approach in geographically distributed sensor systems. *Signal Processing*, **88**, 1698–1714.
- Aziz, Ashraf M. 2014. A new adaptive decentralized soft decision combining rule for distributed sensor systems with data fusion. *Information Sciences*, **256**(Supplement C), 197–210.
- Bai, Yu-Ting, Zhang, Bai-Hai, Wang, Xiao-Yi, Jin, Xue-Bo, Xu, Ji-Ping, Su, Ting-Li, & Wang, Zhao-Yang. 2016. A Novel Group Decision-Making Method Based on Sensor Data and Fuzzy Information. *Sensors (Basel, Switzerland)*, **16**(11), 1799. sensors-16-01799[PII] 27801827[pmid] Sensors (Basel).
- Bak, Per, Tang, Chao, & Wiesenfeld, Kurt. 1987. Self-organized criticality: An explanation of the $1/f$ noise. *Physical Review Letters*, **59**(4), 381–384.
- Banatre, Michel, Marron, Pedro Jose, Ollero, Anibal, & Wolisz, Adam. 2008. *Cooperating Embedded Systems and Wireless Sensor Networks*.

- Barabási, Albert-László; Albert, Réka. 1999. Emergence of scaling in random networks. *Science*, **286**(5439), 509–512.
- Bardossy, Andras, Duckstein, Lucien, & Bogardi, Istvan. 1993. Combination of Fuzzy Numbers Representing Expert Opinions. *Fuzzy Sets Syst.*, **57**(2), 173–181.
- Beheshti, Rahmatollah, Barmaki, Roghayeh, Mozayani, Nasser, & Robu, Valentin. 2016a. *Negotiations in Holonic Multi-agent Systems*. Pages 107–118.
- Beheshti, Rahmatollah, Barmaki, Roghayeh, & Mozayani, Nasser. 2016b. *Negotiations in Holonic Multi-agent Systems*.
- Bellavista, P., Giannelli, C., Lagkas, T., & Sarigiannidis, P. 2018. Multi-domain SDN controller federation in hybrid FiWi-MANET networks. *Eurasip Journal on Wireless Communications and Networking*, **2018**(1).
- Benaskeur, A. R., Irandoust, H., McGuire, P., & Brennan, R. 2007. Holonic control-based Sensor management.
- Berkman, Elliot T., & Reise, Steven P. 2012. *A Conceptual Guide to Statistics Using SPSS*.
- Bernevig, B. A., Giuliano, D., & Laughlin, R. B. 2001. Spinon-Holon Attraction in the Supersymmetric t-J Model with $1/r$. *Phys. Rev. Lett.*, **87**(17).
- Bertossi, Leopoldo E., & Bravo, Loreto. 2004. Query Answering in Peer-to-Peer Data Exchange Systems. *CoRR*, **cs.DB/0401015**.
- Blasch, E., & Plano, S. 2005. *DFIG level 5 (user refinement) issues supporting situational assessment reasoning*.
- Blasch, E., Bossé, É., & Lambert, D. A. 2012a. *High-level Information Fusion Management and Systems Design*.
- Blasch, E., Bossé, É., & Lambert, D. A. 2012b. *High-level Information Fusion Management and Systems Design*.
- Blasch, E., Steinberg, A., Das, S., Llinas, J., Chong, C., O. Kessler, O., Waltz, E., & White, F. 2013 (July). Revisiting the JDL model for information exploitation. *Pages 129–136 of: Proceedings of the 16th International Conference on Information Fusion*.
- Blasch, E. P., Lambert, D. A., Valin, P., Kokar, M. M., Llinas, J., Das, S., Chee, Chong, & Shahbazian, E. 2012c. High Level Information Fusion (HLIF): Survey of models, issues, and grand challenges. *Aerospace and Electronic Systems Magazine, IEEE*, **27**(9), 4–20.
- Bo, C. M., Li, J., Sun, C. Y., & Wang, Y. R. 2003. The application of neural network soft sensor technology to an advanced control system of distillation operation. *International Joint Conference on Neural Networks*, **2**, 1054–1058.
- Boardman, John, & Sauser, Brian. 2008. *Systems thinking. Coping with 21st century problems*.

- Bohanec, Marko, & Rajkovic, Vladislav. 1990. *DEX: An Expert System Shell for Decision Support*. Vol. 1.
- Bossé, É., & Solaiman, B. 2016. *Information Fusion and Analytics for Big Data and IoT*.
- Bossé, Éloi, Roy, Jean, & Wark, Steve. 2007. *Concepts, Models, and Tools for Information Fusion*.
- Boström, Henrik, Andler, Sten, Brohede, Marcus, Johansson, Ronnie, Karlsson, Alexander, Laere, Joeri, Niklasson, Lars, Nilsson, Maria, Persson, Anne, & Ziemke, Tom. 2018. *On the definition of information fusion as a field of research*.
- Botti, Vicente, & Giret, Adriana. 2008. *ANEMONA: A Multi-agent Methodology for Holonic Manufacturing Systems*.
- Bouhouche, Salah, Yah, Mostefa, Hocine, Benjam, & Bast, Jurgen. 2008. Soft sensor - based artificial neural networks and fuzzy logic: application to quality monitoring in hot rolling. *Proceedings of the 10th WSEAS International Conference on Automatic Control, Modelling and Simulation*, 149–154.
- Bourjot, C., Chevrier, V., Di Marzo, G., & Gleizes, M. 2005. Self-Organization in Multi-Agent Systems.
- Brayner, Angelo, Coelho, André L. V., Marinho, Karina, Holanda, Raimir, & Castro, Wagner. 2014. On query processing in wireless sensor networks using classes of quality of queries. *Inf. Fusion*, **15**, 44–55.
- Brehmer, Berndt. 2005. The Dynamic OODA Loop: Amalgamating Boyd’s OODA Loop and the Cybernetic Approach to Command and Control. *Proceedings of the 10th International Command and Control Research Technology Symposium*.
- Bujari, Armir, Conti, Mauro, De Francesco, Carla, & Palazzi, Claudio E. 2019. Fast multi-hop broadcast of alert messages in VANETs: An analytical model. *Ad Hoc Networks*, **82**, 126–133.
- Cameron, C. D., Patsios, C., & Taylor, P. 2015. On the benefits of using self-organising Multi-Agent architectures in network management. *2015 International Symposium on Smart Electric Distribution Systems and Technologies (EDST)*, 335–340.
- Cao, Yue, & Sun, Zhili. 2013. Routing in Delay/Disruption Tolerant Networks: A Taxonomy, Survey and Challenges. *Communications Surveys Tutorials, IEEE*, **15**(2), 654–677.
- Capra, F. 1983. *The Turning Point: Science, Society, and the Rising Culture*.
- Capra, F. 1996. *The Web of Life: A New Scientific Understanding of Living Systems*.
- Castanedo, Federico. 2013. A Review of Data Fusion Techniques. *The Scientific World Journal*, **2013**.

- Cerpa, A., & Estrin, D. 2004. ASCENT: adaptive self-configuring sensor networks topologies. *IEEE Transactions on Mobile Computing*, **3**(3), 272–285.
- Chalkiadakis, Georgios, Elkind, Edith, & Wooldridge, Michael. 2011. *Computational Aspects of Cooperative Game Theory (Synthesis Lectures on Artificial Intelligence and Machine Learning)*.
- Challa, S., Gulrez, T., Chaczko, Z., & Paranesha, T. N. 2005. Opportunistic information fusion: A new paradigm for next generation networked sensing systems. *International Conference on Information Fusion*.
- Chao, Chih-Min, & Hsiao, Tzu-Ying. 2014. Design of structure-free and energy-balanced data aggregation in wireless sensor networks **37**, 229–239.
- Chen, Edwin H. 1971. The Power of the Shapiro-Wilk W Test for Normality in Samples from Contaminated Normal Distributions. *Journal of the American Statistical Association*, **66**(336), 760–762. 01621459.
- Chen, Min, Gonzalez, Sergio, Vasilakos, Athanasios, Cao, Huasong, & Leung, Victor C. 2011. Body Area Networks: A Survey. *Mob. Netw. Appl.*, **16**(2), 171–193.
- Cheng, V. W., & Wang, T. Y. 2012. Performance Analysis of Distributed Decision Fusion Using A Multilevel Censoring Scheme in Wireless Sensor Networks. *IEEE Transactions on Vehicular Technology*, **61**(4), 1610–1619.
- Chiang, S., Huang, C., & Chang, K. 2007. A Minimum Hop Routing Protocol for Home Security Systems Using Wireless Sensor Networks. *IEEE Transactions on Consumer Electronics*, **53**(4), 1483–1489.
- Chirn, Jin-Lung, & McFarlane, Duncan. 2005. Evaluating holonic control systems: a case study. *IFAC Proceedings Volumes*, **38**(1), 211–216.
- Clarke, Ian, Sandberg, Oskar, Wiley, Brandon, & Hong, Theodore W. 2001. *Freenet: A Distributed Anonymous Information Storage and Retrieval System*.
- Clauset, Aaron, Rohilla Shalizi, Cosma, & Newman, M. E. J. 2007. Power-law distributions in empirical data. *SIAM Review*, 661–703.
- Cochez, Michael. 2012. Semantic agent programming language: use and formalization.
- Curiel, Imma. 1997. *Cooperative Game Theory and Applications. Cooperative Games Arising from Combinatorial Optimization Problems*.
- Cybenko, G. 1989. Approximation by superpositions of a sigmoidal function. *Math. Control Signal Systems*, **2**(203), 303–314.
- Dai, Chao, Pang, Jianmin, Zhang, Xiaochuan, Liang, Guanghui, & Bai, Hong. 2016. A novel information fusion model for assessment of malware threat. *International Journal of Security and its Applications*, **10**(5), 1–16.
- Das, S. 2013. *Computational Business Analytics*.

- Das, Subrata. 2008. *High-Level Data Fusion*.
- David de la Peña Esteban, F., A. Lara Torralbo, Juan, Lizcano Casas, David, Aurora Martínez Rey, M., & Pazos Sierra, Juan. 2018. A new approach to computing using informons and holons: towards a theory of computing science. *International Conference on Data Science, E-learning and Information Systems 2018 (Data'18)*.
- De Assis, Adilson José, & Filho, Rubens Maciel. 2000. Soft sensors development for on-line bioreactor state estimation. *Computers & Chemical Engineering*, **24**(2), 1099–1103.
- Delgado, M., Vila, M. A., & Voxman, W. 1998. A fuzziness measure for fuzzy numbers: Applications. *Fuzzy Sets and Systems*, **94**(2), 205–216.
- DeLone, W. H., & McLean, E. R. 1992. Information systems success: The quest for the dependent variable. *Information Systems Research*, 60–95.
- Dervin, Brenda. 1983. An overview of sense-making research: concepts, methods and results to date. *International Communication Association Annual Meeting*.
- Dervin, Brenda. 1998. Sense-making theory and practice: an overview of user interests in knowledge seeking and use. *J. Knowledge Management*, **2**(2), 36–46. . Thu, 23 Oct 2014 17:04:12 +0200.
- Dervin, Brenda. 2003. *Audience as listener and learner, teacher and confidante: The sense-making approach*.
- Devogalaere, Dirk, Rijckaert, Marcel, Leon, Osvaldo Goza, & Lemus, Gil Cruz. 2002. Application of Feedforward Neural Networks for Soft Sensors in the Sugar Industry. *Pages 2– of: Proceedings of the VII Brazilian Symposium on Neural Networks (SBRN'02)*. SBRN '02. Washington, DC, USA: IEEE Computer Society.
- Di Marzo Serugendo, Giovanna, Gleizes, Marie-Pierre, & Karageorgos, Anthony. 2005. Self-organization in multi-agent systems. *Knowl. Eng. Rev.*, **20**(2), 165–189.
- Dietzel, Stefan, Bako, Boto, Schoch, Elmar, & Kargl, Frank. 2009. A fuzzy logic based approach for structure-free aggregation in vehicular ad-hoc networks. *Proceedings of the sixth ACM international workshop on VehiculAr InterNETworking*, 79–88.
- Doksum, Kjell A., & Wong, Chi-Wing. 1983. Statistical Tests Based on Transformed Data. *Journal of the American Statistical Association*, **78**(382), 411–417. 01621459.
- Ebanks, Bruce R. 1983. On measures of fuzziness and their representations. *Journal of Mathematical Analysis and Applications*, **94**(1), 24–37. 0022-247X Added: 2015.10.27.
- Edwards, Mark. 2018. A brief history of holons.
- Ekel, Petr, Kokshenev, Illya, Parreiras, Roberta, Pedrycz, Witold, & Pereira Jr, Joel. 2016. Multiobjective and Multiattribute Decision Making in a Fuzzy Environment and Their Power Engineering Applications. *Inf. Sci.*, **361**(C), 100–119. 0020-0255.

- Eppler, Martin, & Muenzenmayer, Peter. 2002. Measuring Information Quality in the Web Context: A Survey of State-of-the-Art Instruments and an Application Methodology. *Proceedings of the Seventh International Conference on Information Quality (ICIQ-02)*.
- Evans, Martin. 1998. From 1086 and 1984: direct marketing into the millennium. *Marketing Intelligence & Planning*, **16**(1), 56–67.
- F. Buford, John, Jakobson, Gabriel, & Lewis, Lundy. 2010. *Peer-to-peer coupled agent systems for distributed situation management*. Vol. 11.
- Fagin, Ronald, & Halpern, Joseph Y. 1994. Reasoning about knowledge and probability. *J. of the ACM (JACM)*, **41**(2), 340–367.
- Fan, K., Liu, S., & Sinha, P. 2007. Structure-Free Data Aggregation in Sensor Networks. *IEEE Transactions on Mobile Computing*, **6**(8), 929–942.
- Fine, Terrence L. 1977. Review: Glenn Shafer, A mathematical theory of evidence. *Bull. Amer. Math. Soc.*, **83**(4), 667–672.
- Fischer, Klaus, Schillo, Michael, & Siekmann, Jorg. 2003. *Holonic Multiagent Systems: A Foundation for the Organisation of Multiagent Systems*. Vol. 2744. Added: 2015.10.23. Pages 71–80.
- Florentin Smarandache, J. D. 2015. *Advances and Applications of DSMT for Information Fusion, Vol. IV: Collected Works*.
- Flores, Dora-Luz, Rodríguez-Díaz, Antonio, Castro, Juan R., & Gaxiola, Carelia. 2009. *TA-Fuzzy Semantic Networks for Interaction Representation in Social Simulation*.
- Foo, Pek Hui, & Ng, Gee Wah. 2013. High-level Information Fusion: An Overview. *J. of Adv. in Inf. Fusion*, **8**(1).
- Fortuna, L., Graziani, S., & Xibilia, M. G. 2005. Soft sensors for product quality monitoring in debutanizer distillation columns. *Control Engineering Practice*, **13**(4), 499–508.
- Frank, AndrewU. 2008. *Information Processes Produce Imperfections in Data - The Information Infrastructure Compensates for Them*. Pages 467–485.
- Frankel, Justin, & Pepper, Tom. 2000. *The Gnutella Protocol Specification v0.4 (Document Revision 1.2)*. Added: 2015.11.02.
- Gad, A., & Farooq, M. 2002. Data fusion architecture for Maritime Surveillance. *Fifth International Conference on Information Fusion*, **1**, 448–455. . 2015.10.05.
- Gadeo-Martos, Manuel Angel, Fernandez-Prieto, Jose Angel, Canada-Bago, Joaquin, & Velasco, Juan Ramon. 2011. An Architecture for Performance Optimization in a Collaborative Knowledge-Based Approach for Wireless Sensor Networks. *Sensors (Basel, Switzerland)*, **11**(10), 9136–9159. sensors-11-09136[PII] 22163687[pmid] Sensors (Basel).

- Gao, Zhongke, Dang, Weidong, Mu, Chaoxu, Yang, Yuxuan, Li, Shan, & Grebogi, Celso. 2017. A Novel Multiplex Network-based Sensor Information Fusion Model and Its Application to Industrial Multiphase Flow System. *IEEE Transactions on Industrial Informatics*, 1–1.
- García-Marín, A., Jiménez-Hornero, F., & Ayuso, J. L. 2008. La criticalidad autoorganizada y el análisis de datos históricos de lluvia en Córdoba (Andalucía). *Ingeniería del agua*, **Vol. 15, Núm. 1 (2008)**.
- Gerber, Christian, Siekmann, Jorg, & Vierke, Gero. 1999. *Holonic Multi-Agent Systems*.
- Gonsalves, P. G., Cunningham, R., Ton, N., & Okon, D. 2000. Intelligent threat assessment processor (ITAP) using genetic algorithms and fuzzy logic. *Information Fusion*.
- Gonzaga, J. C. B., Meleiro, L. A. C., Kiang, C., & Maciel Filho, R. 2009. ANN-based soft-sensor for real-time process monitoring and control of an industrial polymerization process. *Computers & Chemical Engineering*, **33**(1), 43–49.
- González, Maria Jesús Pidal. 2009. La teoría del caos en las organizaciones. *Cuadernos Unimetanos*, 29–33.
- Groen, Frans C.A., Pavlin, Gregor, Winterboer, Andi, & Evers, Vanessa. 2017. A Hybrid Approach to Decision Making and Information Fusion. *Robot. Auton. Syst.*, **90**(C), 71–85.
- Guo, Y., Yin, C., Li, M., Ren, X., & Liu, P. 2018. Mobile e-Commerce Recommendation System Based on Multi-Source Information Fusion for Sustainable e-Business. *Sustainability*, **10**(1).
- Habib, Carol, Makhoul, Abdallah, Darazi, Rony, & Couturier, Raphaël. 2018. *Health Risk Assessment and Decision-Making for Patient Monitoring and Decision-support using Wireless Body Sensor Networks*. Vol. 47.
- Haken, Hermann, & Portugali, Juval. 2017. Information and Self-Organization. *Entropy* **19**(1), **19**(1).
- Hall, D. L., & Jordan, J. M. 2010. *Human-Centered Information Fusion*.
- Hall, D. L., & Llinas, J. 2001. *Handbook of Multisensor Data Fusion*.
- Hall, David L., & McMullen, Sonya A. H. 2004. *Mathematical Techniques in Multisensor Data Fusion (Artech House Information Warfare Library)*.
- Hall, Kenwood H., Staron, Raymond J., & Vrba, Pavel. 2005. Experience with Holonic and Agent-Based Control Systems and Their Adoption by Industry. 1–10.
- Halpern, Joseph Y. 2003. *Reasoning about uncertainty*.
- Hannah, P., A. Ball, A., & Starr, A. 2000. Decisions in Condition Monitoring - An Exemplar for data fusion Architecture. *International Conference on Information Fusion*.

- Hariharan, S., Bisdikian, C., Kaplan, L.M., & Pham.T. 2011. QoI-based resource allocation for multi-target tracking in energy constrained sensor networks. *14th International Conference on Information Fusion*, 1–8.
- Hassanzadeh, Amin, & Stoleru, Radu. 2014. On the Attack-and-Fault Tolerance of Intrusion Detection Systems in Wireless Mesh Networks. *Pages 387–401 of: Guo, Song, Lloret, Jaime, Manzoni, Pietro, & Ruehrup, Stefan (eds), Ad-hoc, Mobile, and Wireless Networks*. Cham: Springer International Publishing.
- Healing, Alex, Ghanea-Hercock, Robert, Duman, Hakan, & Jakob, Michal. 2008. Nexus: Self-organising Agent-based Peer-to-Peer Middleware for Battlespace Support. *Defense Industry Applications of Autonomous Agents and Multi-Agent Systems*, 1.
- Heinzelman, Bulent Tavli Wendi. 2006. *Mobile Ad Hoc Networks. Energy-Efficient Real-Time Data Communications*.
- Hermans, F., Dziengel, N., & Schiller, J. 2009. Quality estimation based data fusion in wireless sensor networks. *2009 IEEE 6th International Conference on Mobile Adhoc and Sensor Systems*, 1068–1070.
- Hershberger, Scott L. 2011. *Hodges-Lehmann Estimators*.
- Hock, D. W. 1994. *Institutions in the age of mindcrafting*.
- Horling, Bryan, & Lesser, Victor. 2004. A survey of multi-agent organizational paradigms **19**(4), 281–316.
- Hornik, Kurt. 1991. Approximation capabilities of multilayer feedforward networks. *Neural Networks*, **4**(2), 251–257.
- Hsu, Hsi-Mei, & Chen, Chen-Tung. 1996. Aggregation of Fuzzy Opinions Under Group Decision Making. *Fuzzy Sets Syst.*, **79**(3), 279–285.
- Huebscher, C., & McCann, A. 2005. An adaptive middleware framework for context-aware applications. *Personal Ubiquitous Comput.*, **10**(1), 12–20.
- Huebscher, Markus C., & McCann, Julie A. 2004. Adaptive middleware for context-aware applications in smart-homes. *Proceedings of the 2nd workshop on Middleware for pervasive and ad-hoc computing*, 111–116.
- Intanagonwiwat, Chalermek, Govindan, Ramesh, & Estrin, Deborah. 2000. Directed diffusion: a scalable and robust communication paradigm for sensor networks. *Proceedings of the 6th annual international conference on Mobile computing and networking*, 56–67.
- ISO. 2005. *Norma ISO/IEC 20000-2:2005*. . 2015.10.05.
- Iwamura, K., Okubo, N., Tanimizu, Y., & Sugimura, N. 2006. International Journal of Production Research. *International Journal of Production Research*, **44**(18-19), 3657–3675.

- Jakob, Michal, Kaveh, Nima, & Ghanea-Hercock, Robert. 2006. Nexus - Middleware for Decentralized Service-Oriented Information Fusion.
- Janssen, Marijn, van der Voort, Haiko, & Wahyudi, Agung. 2017. Factors influencing big data decision-making quality. *Journal of Business Research*, **70**(C), 338–345.
- Jarimo, Toni, & Salo, Ahti. 2009. Multicriteria partner selection in virtual organizations with transportation costs and other network interdependencies. *Trans. Sys. Man Cyber Part C*, **39**(1), 124–129. . 2015.10.05.
- Jiagao, Wu, Yue, Ma, & Linfeng, Liu. 2018. Cost-efficient dynamic quotacontrolled routing in multi-community delay-tolerant networks. *International Journal of Distributed Sensor Networks*, **14**(5), 1–13.
- Kadlec, P., & Gabrys, B. 2007. Nature-inspired Adaptive Architecture for Soft Sensor Modelling. *NiSIS'2007 Symposium: 3rd European Symposium on Nature-inspired Smart Information Systems*.
- Kamarei, Meisam, Hajimohammadi, Mojtaba, Patooghy, Ahmad, & Fazeli, Mahdi. 2015. An Efficient Data Aggregation Method for Event-Driven WSNs: A Modeling and Evaluation Approach. *Wirel. Pers. Commun.*, **84**(1), 745–764.
- Kandfer, Michael, & Shapiro, Stuart. 2008. *A Categorization of Contextual Constraints*.
- Karageorgos, Giovanna Di Marzo; Marie-Pierre Gleizes;Anthony. 2011. *Self-Organising Systems*. Pages 7–32.
- Karagiannis, G., Altintas, O., Ekici, E., Heijenk, G., Jarupan, B., Lin, K., & Weil, T. 2011. Vehicular Networking: A Survey and Tutorial on Requirements, Architectures, Challenges, Standards and Solutions. *Communications Surveys Tutorials, IEEE*, **13**(4), 584–616.
- Kessel, R. T. 2006. The Dynamics of Information Fusion: Synthesis Versus Misassociation. *2006 9th International Conference on Information Fusion*, 1–6.
- Khaleghi, Bahador, Khamis, Alaa, & Karray, Fakhri. 2015. *Multisensor Data Fusion a Data Centric Review of the State-of-the-Art and Overview of Emerging Trends*.
- Kim, Hyun Ho, Jo, Hyeong Gon, & Kang, Soon Ju. 2017. Self-Organizing Peer-To-Peer Middleware for Healthcare Monitoring in Real-Time. *Sensors (Basel, Switzerland)*, **17**(11), 2650.
- Knight, Shirlee-ann, & Burn, Janice. 2005. Developing a Framework for Assessing Information Quality on the World Wide Web. *Informing Science*.
- Koestler, Arthur. 1982. *The ghost in the machine*. . 2015.10.05.
- Koitzsch, A., Borisenko, S. V., Geck, J., Zabolotnyy, V. B., Knupfer, M., J. Fink, P. Ribeiro, Buechner, B., & Follath, R. 2006. Current spinon-holon description of the one-dimensional charge-transfer insulator SrCuO₂: Angle-resolved photoemission measurements. *Physics Review*, **B 73**.

- Kokar, Mieczyslaw M., Bedworth, Mark D., & Frankel, Carl B. 2000. *Reference model for data fusion systems*.
- Kordon, A., Smits, G., Jordaan, E., & Richtor, E. 2002. Robust soft sensors based on integration of genetic programming, analytical neural networks, and support vector machines. *Congress on Evolutionary Computation, 2002. CEC '02.*, **1**, 896–901.
- Krishna, Vijay. 2003. *Auction Theory*.
- Lai, Y., & Chen, H. 2007. Energy-Efficient Fault-Tolerant Mechanism for Clustered Wireless Sensor Networks. *2007 16th International Conference on Computer Communications and Networks*, 272–277.
- Lambert, Dale A. 2009. A blueprint for higher-level fusion systems. *Inf. Fusion*, **10**(1), 6–24.
- Lan, Yanting, Huang, Jinying, & Chen, Xiaodong. 2016. Application of two-level joint information fusion model in intelligent vehicle. *International Journal of Interactive Mobile Technologies*, **12**(3), 77–81.
- Laureyssens, Dirk. 2009.
- Laxhammar, R., Falkman, G., & Sviestins, E. 2009. Anomaly detection in sea traffic - A comparison of the Gaussian Mixture Model and the Kernel Density Estimator. *2009 12th International Conference on Information Fusion*, 756–763.
- Lee, Joo-Ho, Morioka, K., Ando, N., & Hashimoto, H. 2004. Cooperation of distributed intelligent sensors in intelligent environment. *IEEE/ASME Transactions on Mechatronics*, **9**(3), 535–543.
- Li, Bo, & Pang, Fuwen. 2018. Innovative assessment scheme of navigation risk based on improved multi-source information fusion techniques. *International Journal of Distributed Sensor Networks*, **14**(4), 1550147718772543.
- Lichman, M. 2013. *UCI Machine Learning Repository*.
- Liggins, M., Hall, D., & Llinas, J. 2008. *Handbook of Multisensor Data Fusion: Theory and Practice, Second Edition*.
- Little, Eric G., & Rogova, Galina L. 2009. Designing ontologies for higher level fusion. *Inf. Fusion*, **10**(1), 70–82.
- Llinas, J. 1988. *Toward the Utilization of Certain Elements of AI Technology for Multi Sensor Data Fusion*.
- Maitlis, Sally, & Christianson, Marlys. 2014. Sensemaking in organizations: Taking stock and moving forward. *The Academy of Management Annals*, **8**(1), 57–125.
- Marik, Vladimir, Vyatkin, Valeriy, & Colombo, Armando Walter. 2007. *Holonic and Multi-Agent Systems for Manufacturing*.

- Marik, Vladimir, Martinez Lastra, Jose L., & Skobelev, Petr. 2013. *Industrial Applications of Holonic and Multi-Agent Systems*.
- Martínez Luaces, Milton, Gayoso, Celina, Pazos Sierra, Juan, & Rodríguez-Patón, Alfonso. 2008. Using modular neural networks to model self-consciousness and self-representation for artificial entities. *International Journal of Mathematics and Computers in Simulation*, **2**(2).
- Mastrogiovanni, Fulvio, Sgorbissa, Antonio, & Zaccaria, Renato. 2007. A distributed architecture for symbolic data fusion. *Proceedings of the 20th international joint conference on Artificial intelligence*, 2153–2158.
- Mathews, John. 1996. Holonic organisational architectures. *Hum. Syst. Manag.*, **15**(1), 27–54.
- McFarlane, Duncan C., & Bussmann, Stefan. 2000. Developments in holonic production planning and control. *Production Planning & Control*, **11**(6), 522–536.
- McHugh, P., & Merli, G. W. 1995. *Beyond Business Process Reengineering - Towards the Holonic Enterprise*.
- McKendall, R., & Mintz, M. 1988 (April). Robust fusion of location information. *Pages 1239–1244 vol.2 of: Proceedings. 1988 IEEE International Conference on Robotics and Automation*.
- Meer, Hermann De, & Koppen, Christian. 2005a. *Characterization of Self-Organization*.
- Meer, Hermann De, & Koppen, Christian. 2005b. *Self-organization in Peer-to-Peer Systems*.
- Mella, Piero. 2012. *Systems Thinking Intelligence in Action*.
- Mella, Piero, & Gazzola, Patrizia. 2017. The Holonic View of Organizations and Firms. *Systems Research and Behavioral Science*, **34**(3), 354–374.
- Milic, Bratislav, & Malek, Mirosław. 2009. *Guide to Wireless Ad Hoc Networks*.
- Miller, Brent, & Steyvers, Mark. 2011. The Wisdom of Crowds with Communication. *In: CogSci*.
- Miller, Holmes. 1996. The multiple dimensions of information quality. *Information Systems Management*, **13**(2), 79–82.
- Mnatsakanyan, Zaruhi R., Burkom, Howard S., Hashemian, Mohammad R., & Coletta, Michael A. 2012. Distributed information fusion models for regional public health surveillance. *Information Fusion*, **13**(2), 129–136.
- Mönks, Uwe. 2017. *Information Fusion Under Consideration of Conflicting Input Signals*.

- Mozaffari, Ahmad, & Azad, Nasser L. 2015. An ensemble neuro-fuzzy radial basis network with self-adaptive swarm based supervisor and negative correlation for modeling automotive engine coldstart hydrocarbon emissions. *Appl. Soft Comput.*, **32**(C), 449–467.
- Mujeeb Ur, Rehman, Sheeraz, Ahmed, Sarmad Ullah, Khan, Shabana, Begum, & Ahmed, S. Hassan. 2018. Performance and Execution Evaluation of VANETs Routing Protocols in different Scenarios. *EAI Endorsed Transactions on Energy Web, Vol 5, Iss 17, Pp 1-5 (2018)*, 1.
- Nakamura, Eduardo Freire, Loureiro, Antonio Alfredo Ferreira, Boukerche, Azzedine, & Zomaya, Albert Y. 2014. Localized algorithms for information fusion in resource constrained networks. *Inf. Fusion*, **15**, 2–4.
- Napoli, G., & Xibilia, M. G. 2011. Soft Sensor design for a Topping process in the case of small datasets. *Computers & Chemical Engineering*, **35**(11), 2447–2456.
- Napster, www.napster.com.
- Naumann, Felix, & Rolker, Claudia. 2000. *Assessment Methods for Information Quality Criteria*.
- Ngai, Edith, Gelenbe, Erol, & Humber, Gregory. 2009. *Information-Aware Traffic Reduction for Wireless Sensor Networks*.
- Ngai, Edith C. H., & Gunningberg, Per. 2014. Quality-of-information-aware data collection for mobile sensor networks. *Pervasive and Mobile Computing*, **11**, 203–215.
- Niu, Ruixin, Chen, Biao, & Varshney, P.K. 2006. Fusion of decisions transmitted over Rayleigh fading channels in wireless sensor networks. *Trans. Sig. Proc.*, **54**(3), 1018–1027.
- Novák, V. 2005. Are Fuzzy Sets a Reasonable Tool for Modeling Vague Phenomena? *Fuzzy Sets and Systems*, **156**., 341–348.
- O'Reilly, C. A. 1982. Variations in decision makers' use of information sources: The impact of quality and accessibility of information. *Academy of Management Journal*, 756–771.
- Ou, Zhonghong, Harjula, Erkki, Kassinen, Otso, & Ylianttila, Mika. 2010. Performance evaluation of a Kademia-based communication-oriented P2P system under churn. *Comput. Netw.*, **54**(5), 689–705.
- Ozdemir, Suat, & Xiao, Yang. 2013. FTDA: outlier detection-based fault-tolerant data aggregation for wireless sensor networks. *Security and Communication Networks*, **6**(6), 702–710.
- Padgham, Lin, & Winikoff, Michael. 2005. *Appendix C: The AUMN Notation*.
- Paggi, H., & Alonso, F. 2010. *Uncertainty and Randomness: A Holonic Approach*.

- Paggi, H., Bossé, E., Florea, M. C., & Solaiman, B. 2014. *On the use of holonic agents in the design of information fusion systems*.
- Paggi Straneo, Horacio, Alonso Amo, Fernando, & Crespo del Arco, Jose. 2010. Cohesion and Complex System Structure. *Systemics and Informatics World Network SIWN 2010*.
- Pani, Ajaya Kumar, Amin, Krunal G., & Mohanta, Hare Krishna. 2016. Soft sensing of product quality in the debutanizer column with principal component analysis and feed-forward artificial neural network. *Alexandria Engineering Journal*, **55**(2), 1667–1674.
- Pichler, Franz. 2000. On the Construction of A. Koestler's Holarchical networks. *European Meetings on Cybernetics and Systems Research (EMCSR)*.
- Price, Rosanne; Shanks, Graeme. 2011. The Impact of Data Quality Tags on Decision-Making Outcomes and Process. *Journal of the Association for Information System*, **12**(4), 323–346.
- Pu, Qiumei, Lin, Ouyang, & Fu, S. 2011. *Adopting semantic language in agent communication processes*.
- Rafique, Irfan, Lew, Philip, Abbasi, Maissom Qanber, & Li, Zhang. 2012. Information Quality Evaluation Framework: Extending ISO 25012 Data Quality Model. *World Academy of Science, Engineering and Technology*, **6**(5).
- Rani, Asha, Singh, Vijander, & Gupta, J. R. P. 2013. Development of soft sensor for neural network based control of distillation column. *ISA transactions*, **52**(3), 438–449.
- Raol, Jitendra R. 2015. *Data Fusion Mathematics. Theory and Practice*.
- Richardson, J. M., & Marsh, K. A. 1988. Fusion of Multisensor data. *The International Journal of Robotics Research*, **7**(6), 78–96.
- Rogova, G., Hadzagic, M., St-Hilaire, M., Florea, M. C., & Valin, P. 2013. Context-based information quality for sequential decision making. *2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*, 16–21.
- Rowstron, Antony, & Druschel, Peter. 2001. Pastry: Scalable, Decentralized Object Location, and Routing for Large-Scale Peer-to-Peer Systems. 329–350.
- Russell, Daniel M., Stefik, Mark J., Pirolli, Peter, & Card, Stuart K. 1993. The cost structure of sensemaking. *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, 269–276.
- Sakia, R. M. 1992. The Box-Cox Transformation Technique: A Review. *Journal of the Royal Statistical Society. Series D (The Statistician)*, **41**(2), 169–178. 00390526, 14679884.
- Samir Dani, Backhouse C.J., Burns N.D. 2004. Application of Transactional Analysis in Supply Chain Networks: A Potential Holonic Mediating Tool. *Proceedings of the Institution of Mechanical Engineers. Part B: Journal of Engineering Manufacture*, **5**(218), 571–580.

- Saporo, Agus. 2014. State of the Art in the Development of Adaptive Soft Sensors based on Just-in-Time Models. *Procedia Chemistry*, **9**(Supplement C), 226–234.
- Sasaki, T., Brscic, D., & Hashimoto, H. 2007. Implementation of Distributed Sensor Network for Intelligent Space. *2007 IEEE International Conference on Mechatronics*, 1–6.
- Schillo, Michael, & Fischer, Klaus. 2004. *A Taxonomy of Autonomy in Multiagent Organisation*.
- Schuldt, Arne. 2011. *Multiagent Coordination Enabling Autonomous Logistics*.
- Schwaiger, Arndt, & Stahmer, Björn. 2005. *Probabilistic Holons for Efficient Agent-Based Data Mining and Simulation*.
- Shafaei, S., & Aghaee, N. G. 2008. Biological Network Simulation Using Holonic Multi-agent Systems. *Tenth International Conference on Computer Modeling and Simulation (uksim 2008)*, 617–622.
- Shimizu, H. 1987. *A General Approach to Complex Systems in Bioholonics*. Pages 204–223.
- Simon, H. A. 1957. *Models of a man*.
- Simon, H. A. 1996. *The Sciences of the Artificial*.
- Singh, Alka, Rathkanthiwar, Shubhangi, & Kakde, Sandeep. 2016. *LEACH based-energy efficient routing protocol for wireless sensor networks*.
- Singh, Suresh, & Raghavendra, C. S. 1998. PAMAS—power aware multi-access protocol with signalling for ad hoc networks. *SIGCOMM Comput. Commun. Rev.*, **28**(3), 5–26.
- Smarandache, Florentin, & Dezert, Jean. 2004. The Combination of Paradoxical, Uncertain, and Imprecise Sources of Information based on DSMT and Neutro-Fuzzy Inference. *CoRR*, **abs/cs/0412091**. . Wed, 10 Oct 2012 21:28:56 +0200.
- Smith, N. J. J. 2011. *Vagueness, Uncertainty and Degrees of Belief: Two Kinds of Indeterminacy; One Kind of Credence*.
- Smuts, J. Ch. 1926. *Holism and evolution*.
- Snidaro, Lauro, García, Jesús, & Llinas, James. 2015. Context-based Information Fusion: A survey and discussion. *Information Fusion*, **25**, 16–31.
- Soares, Symone G., Takao, Thales Baliero, Rocha, A. F. Da, de A. Araújo, Ricardo, & de A. Barbosa, Talles M. G. 2011. Building Distributed Soft Sensors. *International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM)*, **3**, 202–209.

- Solaiman, B., Bossé, É., Pigeon, L., Guériot, D., & Florea, M. C. 2015. A conceptual definition of a holonic processing framework to support the design of information fusion systems. *Inf. Fusion*, **21**, 85–99.
- Solano-Aragón, Cinthya, & Garza, Arnulfo Alanis. 2009. *Multi-Agent System with Fuzzy Logic Control for Autonomous Mobile Robots in Known Environments*. . Thu, 22 Apr 2010 20:03:12 +0200. Pages 33–52.
- Soriano, Francisco Javier. 2003. Modelo de Arquitectura para Gestion Cooperativa de Sistemas y Servicios Distribuidos Basado en Agentes Autonomos.
- Starnes, Daren S., Yates, Dan, & Moore, David S. 2014. *The Practice of Statistics*.
- Steinberg, Alan N., Bowman, Christopher L., & White, Franklin E. 1999. *Revisions to the JDL data fusion model*.
- Stutzbach, Daniel, & Rejaie, Reza. 2006. *Understanding churn in peer-to-peer networks*. . 2015.10.05.
- Sulis, William, Nation, James, Trofimova, Irina, Rand, John, & Sulis, William. 2003. *Archetypal Dynamics*.
- Sutton, Peter. 2013. Vagueness, Communication, and Semantic Information. . 2015.10.05.
- Tavli, Bulent, & Heinzelman, Wendi. 2006. *Mobile Ad Hoc Networks. Energy-Efficient Real-Time Data Communications*.
- Thiéart, R. A., & Forgues, B. 1995. Chaos Theory and Organization. *Organization Science*, **6**(1), 19–31.
- Turnbull, S. 2001. *Grounding the theory of the firm in the natural sciences*. . 2015.10.05.
- Turnbull, Shann. 2000. The Governance of Firms Controlled By More Than One Board: Theory Development And Examples. . 2015.10.05.
- Turnbull, Shann. 2003. *Emergence of a Global Brain: For and from World Governance*. . 2015.10.05.
- Tutschku, Kurt, & Tran-Gia, Phuoc. 2005. 23. *Traffic Characteristics and Performance Evaluation of Peer-to-Peer Systems*.
- Ulieru, Mihaela. 2002a. *Emergence of Holonic Enterprises from Multi-Agent Systems: A Fuzzy-Evolutionary Approach*.
- Ulieru, Mikaela. 2002b. A Fuzzy Mathematics Approach to Modeling Emergent Holonic Structures. *Pages 241–255 of: Cleja-Tigoiu, Sanda, & Tigoiu, Victor (eds), Geometry, Continua & Microstructures*. Academic Press.
- Unland, Rainer, & Ulieru, Mihaela. 2005. Swarm Intelligence and the Holonic Paradigm: A Promising Symbiosis. *Knowledge-Based Intelligent Information and Engineering Systems: 9th International Conference, KES 2005*.

- V K, Subhashree, & Tharini, C. 2017. *An Energy Efficient Routing and Fault Tolerant Data Aggregation (EERFTDA) algorithm for wireless sensor networks*. Vol. 23.
- Valckenaers, Paul, & Van Brussel, Hendrik. 2005. *Fundamental Insights into Holonic Systems Design*.
- Valencia-Jiménez, J. J., & Fernández-Caballero, Antonio. 2008. *Holonic Multi-agent System Model for Fuzzy Automatic Speech / Speaker Recognition*. 978-3-540-78582-8. Pages 73–82.
- Villas, L. A., Boukerche, A., Ramos, H. S., de Oliveira, H. A. B. F., de Araujo, R. B., & Loureiro, A. A. F. 2013. DRINA: A Lightweight and Reliable Routing Approach for In-Network Aggregation in Wireless Sensor Networks. *IEEE Transactions on Computers*, **62**(4), 676–689.
- Vu, Quang, Hieu, Lupu, Mihai, Ooi, & Beng, Chin. 2010a. *Architecture of Peer-to-Peer Systems*.
- Vu, Quang Hieu, Lupu, Mihai, & Ooi, Beng Chin. 2010b. *Peer-to-Peer Computing. Principles and Applications*.
- Wald, L. 1999. Some terms of reference in data fusion. *IEEE Transactions on Geoscience and Remote Sensing*, **37**(3), 1190–1193.
- Waltz, E. L., & Llinas, J. 1990. *Multisensor Data Fusion*.
- Wand, Yair, & Wang, Richard Y. 1996. Anchoring data quality dimensions in ontological foundations. *Commun. ACM*, **39**(11), 86–95.
- Wang, Jie-Sheng, & Han, Shuang. 2015. Feed-Forward Neural Network Soft-sensor Modeling of Flotation Process Based on Particle Swarm Optimization and Gravitational Search Algorithm. *Intell. Neuroscience*, **2015**(Jan.), 1:1–1:1.
- Wang, X., & Hovakimyan, N. 2010. L1-adaptation-based distributed event-triggering. *49th IEEE Conference on Decision and Control (CDC)*, 7570–7575.
- Wang, X., & Hovakimyan, N. 2012. *Performance prediction in uncertain systems using L1 adaption-based distributed events triggering*. Vol. 1.
- Watts, Duncan J., & Strogatz, Steven H. 1998. Collective dynamics of 'small-world' networks. *Nature*, 393–440.
- Wee Siong, Ng, Beng Chin, Ooi, & Kian-Lee, Tan. 2002. BestPeer: a self-configurable peer-to-peer system. *Proceedings 18th International Conference on Data Engineering*, 272.
- Weick, Karl E. 1979. *The Social Psychology of Organizing*.
- Weick, Karl E. 1993. The Collapse of Sensemaking in Organizations: The Mann Gulch Disaster. *Administrative Science Quarterly*, **38**, 628–652.

- Weick, Karl E. 1995. *Sensemaking in organizations*.
- Wilber, Ken. 2000. *A Brief History of Everything*.
- Wilber, Ken. 2007. *The Integral Vision: A Very Short Introduction to the Revolutionary Integral Approach to Life, God, the Universe, and Everything*.
- Witte, Robert S., & Witte, John S. 2013. *Statistics*.
- Ye, Wei, Heidemann, J., & Estrin, D. 2002. *An energy-efficient MAC protocol for wireless sensor networks*.
- Yousefi, Hamed, Yeganeh, Mohammad Hossein, Alinaghpour, Naser, & Movaghar, Ali. 2012. Structure-free real-time data aggregation in wireless sensor networks. *Comput. Commun.*, **35**(9), 1132–1140.
- Zadeh, L. 1973. Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Trans. SMC*, **3**(1), 28–44.
- Zeleznikar, Anton. 2002. Informon- an emergent conscious component. *Informatica*. . 2015.10.05.
- Zhang, Jie, Gao, Liang, Chan, Felix, & Li, Peigen. 2003. *A holonic architecture of the concurrent integrated process*. Vol. 139.
- Zhao, Z. 2009. Data Quality-Oriented Data Integration in Peer-to-Peer System. *2009 Ninth International Conference on Hybrid Intelligent Systems*, **2**, 419–422.
- Zhong, Hong, Shao, Lili, Cui, Jie, & Xu, Yan. 2018. An efficient and secure recoverable data aggregation scheme for heterogeneous wireless sensor networks. *J. Parallel Distrib. Comput.*, **111**(C), 1–12.
- Zhou, Dengji, Wei, Tingting, Zhang, Huisheng, Ma, Shixi, & Wei, Fang. 2018. An information fusion model based on dempster-shafer evidence theory for equipment diagnosis. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering*, **4**(2).
- Zhu, Xiaolan, & Gauch, Susan. 2000. Incorporating Quality Metrics in Centralized/Distributed Information Retrieval on the World Wide Web. *Pages 288–295 of: Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. SIGIR '00. New York, NY, USA: ACM.
- Zi-Xiao, Wang. 1984. Fuzzy measures and measures of fuzziness. *Journal of Mathematical Analysis and Applications*, **104**(2), 589–601. 0022-247X Added: 2015.10.27.

ANEXOS

ANEXO A: DATOS DEL CASO DE ESTUDIO SOBRE NIVELES DE OZONO

Los datos disponibles consisten en registros de 72 columnas, siendo cada una de ellas un parámetro (correspondiente a un campo en nuestro modelo) que determinan si el nivel de ozono es alto o no, por ejemplo temperatura máxima, temperatura promedio, humedad relativa, presión atmosférica a nivel del mar, precipitaciones, radiación total del día, temperaturas y vientos medidos a distintas horas del día. Algunos de esos parámetros pueden considerarse compuestos por otros de la lista (por ejemplo TT depende de T850, T500, Td850 y T500, KI depende de T850, T500, DP850 y DP700). El campo de decisión (de si es un día de ozono alto o no) se considera dependiente de todos los demás.

Todos los atributos que comienzan por T están asociados a la temperatura medida en distintos momentos a través del día y los que empiezan por WS indican la velocidad del viento en distintos momentos.

WSR_PK: continuo. Velocidad máxima del viento

WSR_AV: continuo. Velocidad del viento promedio

T_PK: continuo. T máx.

T_AV: continuo. T promedio

T85: continuo. T a 850 hectopascales (hpa)

RH85: continuo. Humedad relativa a 850 hpa

U85: continuo. (viento dirección este-oeste a 850 hpa)

V85: continuo. Viento dirección N-S a 850 hpa.

HT85: continuo. Altura geopotencial a 850 hpa,(aproximadamente igual a la altura a baja altitud)

T70: continuo. T a nivel correspondiente a 700 hpa (aproximadamente 3100 m de altura)

RH70: continuo. U70: continuo. V70: continuo. HT70: continuo.
 T50: continuo. T a nivel de 500 hpa (aproximadamente a 5500 m de altura)
 RH50: continuo. U50: continuo.
 V50: continuo.
 HT50: continuo.
 KI: continuo.

KI Es una medida del potencial que existe de que se produzca una tormenta basada en las tasas de temperaturas verticales, la extensión vertical de la capa de humedad y el contenido de humedad de la atmósfera inferior. La diferencia de las temperaturas entre 850 mb y 500 mb es usada para parametrizar la tasa de lapsos de temperatura vertical. Los 850 puntos de rocío proveen información del contenido de humedad de la atmósfera baja. La extensión vertical de la capa de humedad es representada por la diferencia de la temperatura de 700 mb y el punto de rocío a 700mb. Esa diferencia es llamada la depresión de rocío a 700 mb. El índice se deriva aritméticamente:

K-index = (temperatura a 850 mb - temperatura a 500 mb) + punto de rocío a 850 mb - depresión del punto de rocío a 700 mb

TT: continuo. El TT (*Total Totals*) es un índice usado para evaluar la fuerza de una tormenta. Es una combinación del total vertical (VT) y los totales cruzados (CT). El VT es la diferencia de temperatura entre 850 y 500 mb mientras que el total cruzado CT es el punto de rocío a 850 mb menos la temperatura a 500 mb:

$$TT = (T_{850} - T_{500}) + (Td_{850} - T_{500})$$

Por ejemplo:

temperatura a 850 mb = 20 grados C

temperatura a 500 mb = -7.5 grados C

punto de rocío a 850 mb = 9 grados C

$$TT = (20 - (-7.5)) + (9 - (-7.5)) = 27.5 + 16.5 = 44$$

El significado de ese TT es mostrado en el la tabla 19.

TT	Significado
<44	Convección poco probable
44-50	Probables tormentas
51-52	Tormentas severas aisladas
53-56	Tormentas severas muy aisladas
>56	Tormentas severas aisladas

Tabla 19: Significado del valor TT

SLP: continuo. Presión a nivel del mar.

SLP_: continuo. Cambio en el SLP respecto del día anterior.

Precp: continuo.—precipitación

Número de instancias (registros): 2536

Número de atributos: 73

Información de los atributos

Clase: 1,0 | Dos clases: 1: día con problemas en el ozono, 0: día normal.

Se ignora la fecha.

Los campos se consideran numerados del 1 al 72, siendo el 1 el WSR0 y el 72 Precp.
El campo 73 es el de resultado (clase), según la lista:

WSR0: continuo.	T9: continuo.
WSR1: continuo.	T10: continuo.
WSR2: continuo.	T11: continuo.
WSR3: continuo.	T12: continuo.
WSR4: continuo.	T13: continuo.
WSR5: continuo.	T14: continuo.
WSR6: continuo.	T15: continuo.
WSR7: continuo.	T16: continuo.
WSR8: continuo.	T17: continuo.
WSR9: continuo.	T18: continuo.
WSR10: continuo.	T19: continuo.
WSR11: continuo.	T20: continuo.
WSR12: continuo.	T21: continuo.
WSR13: continuo.	T22: continuo.
WSR14: continuo.	T23: continuo.
WSR15: continuo.	T_PK: continuo.
WSR16: continuo.	T_AV: continuo.
WSR17: continuo.	T85: continuo.
WSR18: continuo.	RH85: continuo.
WSR19: continuo.	U85: continuo.
WSR20: continuo.	V85: continuo.
WSR21: continuo.	HT85: continuo.
WSR22: continuo.	T70: continuo.
WSR23: continuo.	RH70: continuo.
WSR_PK: continuo.	U70: continuo.
WSR_AV: continuo.	V70: continuo.
T0: continuo.	HT70: continuo.
T1: continuo.	T50: continuo.
T2: continuo.	RH50: continuo.
T3: continuo.	U50: continuo.
T4: continuo.	V50: continuo.
T5: continuo.	HT50: continuo.
T6: continuo.	KI: continuo.
T7: continuo.	TT: continuo.
T8: continuo.	SLP: continuo.
	SLP_: continuo. Precp:
	continuo.

- La figura 31 muestra un grafo con parte de las dependencias entre los campos.

Figura 31: Vista parcial de las relaciones de dependencia del caso A.

Ejemplo de un registro de datos

En el juego de datos original:

1/1/1998,0.8,1.8,2.4,2.1,2.2,1.5,1.7,1.9,2.3,3.7,5.5,5.1,5.4,5.4,4.7,4.3,3.5,3.5,2.9,3.2,3.2,2.8,2.6,5.5,3.1,5.2,6.1,
6.1,6.1,6.1,5.6,5.2,5.4,7.2,10.6,14.5,17.2,18.3,18.9,19.1,18.9,18.3,17.3,16.8,16.1,15.4,14.9,14.8,15,19.1,12.5,
6.7,0.11,3.83,0.14,1612,-2.3,0.3,7.18,0.12,3178.5,-15.5,0.15,10.67,-1.56,5795,-12.1,17.9,10330,-55,0,0.

En el juego de datos usado:

0.8,1.8,2.4,2.1,2.2,1.5,1.7,1.9,2.3,3.7,5.5,5.1,5.4,5.4,4.7,4.3,3.5,3.5,2.9,3.2,3.2,2.8,2.6,5.5,3.1,5.2,
6.1,6.1,6.1,6.1,5.6,5.2,5.4,7.2,10.6,14.5,17.2,18.3,18.9,19.1,18.9,18.3,17.3,16.8,16.1,15.4,14.9,14.8,15,19.1,
12.5,6.7,0.11,3.83,0.14,1612,-2.3,0.3,7.18,0.12,3178.5,-15.5,0.15,10.67,-1.56,5795,-12.1,17.9,10330,-55,0,No

Las dependencias se pueden graficar según la figura 31 y están dadas por una matriz binaria del estilo de la de la tabla 20 (aquí no se la muestra completa por un tema de espacio). Un 1 en la posición i, j significa que el campo i depende en forma directa del campo j .

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	1	1	1	0	0
0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	1	1	1	1	1
0	0	0	0	0	0	0	0	0	0
1	1	1	0	0	0	0	0	0	1

Tabla 20: Modelo de matriz de dependencias de campos

ANEXO B: DATOS DEL CASO DE ESTUDIO DE COMPRA DE VEHÍCULO

La base de datos de evaluación de autos se deriva de un modelo de decisión jerárquica simple desarrollado originalmente para la demostración del marco (*shell*) de sistema experto para la toma de decisiones multi-atributo DEX (Bohanec & Rajkovic, 1990). El modelo evalúa coches de acuerdo a la estructura conceptual descrito por la figura 32 y la tabla 21, donde:

Campo		Formado por		Significado
Car	PRICE	buying		<i>Precio de compra</i>
		maint		<i>Precio de mantenimiento</i>
	TECH			<i>Características técnicas</i>
		COMFORT		<i>Confort</i>
			Doors	<i>Número de puertas</i>
			Persons	<i>Capacidad en términos de personas que puede llevar</i>
			lug. Boot	<i>Tamaño del maletero</i>
		Safety		<i>Seguridad estimada del coche</i>

Cuadro 9.3: Composición de los campos caso B.

Los atributos de entrada (elementales) se indican en minúsculas. Más allá del concepto meta (CAR), el modelo incluye tres conceptos intermedios: PRICE, TECH, COMFORT. Cada concepto está relacionado en el modelo original con sus descendientes en minúscula. **Número de instancias:** 1728 (las instancias cubren completamente el espacio de atributos).

Número de atributos: Seis

Valores de los atributos:

- buying v-high, high, med, low
- maint v-high, high, med, low
- doors 2, 3, 4, 5-more
- persons 2, 4, more
- lug_boot small, med, big
- safety low, med, high

Valores de CAR: unacc, acc, good, vgood

No hay atributos sin valor.

Figura 32: Composición de los campos caso B.

Ejemplo de los registros usados:

Precio compra; Precio manten; PRICE; Nro. puertas; Nro pasajeros ;Baul ;COMFORT;
 Seguridad; TECH;CAR CALCULADO 1;1;1;2;2;1;4;1;4;4

ANEXO C: DATOS DEL CASO DE ESTUDIO DE CLASIFICACIÓN DE SITIOS WEB

El problema del *phishing* es considerado una cuestión vital en la industria, especialmente en banca y comercio electrónico si se consideran el número de transacciones en línea que implican pagos.

Se han identificado distintas características relacionadas con sitios web legítimos y de *phishing* y se recolectaron 1353 nombres de sitios web de distintas fuentes. Los sitios de *phishing* se recolectaron de Phishtank (www.phishtank.com), que es un sitio gratuito donde los usuarios pueden enviar, verificar, rastrear y compartir datos de *phishing*. Los sitios legítimos fueron recolectados de Yahoo usando un programa *ad hoc*. En el juego de datos original hay 702 URLs de *phishing* y 103 sospechosas.

Cuando un sitio web es considerado SOSPECHOSO significa que puede ser tanto de *phishing* como legítimo indicando que el sitio tiene algunas características de uno legítimo y otras de uno de *phishing*.

Hay varios campos compuestos, los que a su vez dependen de los valores de otros campos que por brevedad no se mencionan:

- ADDRESS BAR BASED FEATURES
- ABNORMAL BASED FEATURES
- HTML- AND JAVASCRIPT-BASED FEATURES
- DOMAIN BASED FEATURES

La clasificación del sitio (RESULT) depende de todos los campos compuestos. Por ejemplo ADDRESS BAR BASED FEATURES depende de :

- Uso de la dirección IPs o URL larga para ocultar la parte sospechosa.
- Uso de servicios que abrevian la URL o URL conteniendo “@” y usando “//” .
- Adición de un prefijo o sufijo separado por - al dominio o Sub dominio y múltiples sub-dominios .
- HTTPS (*Hyper Text Transfer Protocol with Secure Sockets Layer*).
- *Domain Registration Length*.
- *Favicon*.
- Uso de puertos no estándares.
- Existencia de la cadena “HTTPS” en la parte de nombre del dominio de la URL.

Los campos son descritos en detalle en (Abdelhamid *et al.*, 2014).

Dos ejemplos de registros son: 1,1,1,1,-1,-1,-1,-1,-1,1,1,-1,1,-1,-1,-1,0,1,1,1,1,-1,-1,-1,-1,1,1,-1, CLASS=-1

1,1,1,1,1,-1,0,1,-1,1,1,-1,1,0,-1,-1,1,1,0,1,1,1,1,-1,-1,0,-1,1,1,1, CLASS= -1

donde -1 significa sitio legítimo , 0 sospechoso y 1 sitio de *phishing* .

Se tiene además un conjunto de reglas basadas en atributos del sitio que dicen si un sitio web es o no de *phising*, las reglas se agrupan por tópico, por ejemplo, características basadas en el nombre del dominio o en su dirección. A su vez, hay datos de distintos sitios con valores para cada atributo.

El resultado final, RESULT (si un sitio web es o no de *phising*) depende de los campos a) b), c) y d), los que se determinan por una serie de reglas.

La tabla 22 describe los valores posibles para cada campo y la figura 33 da un esquema parcial de como se componen los mismos.

Nombres de campo y {valores posibles}		
<i>having_IP_Address</i>	<i>URL_Length</i>	<i>Shortining_Service</i>
{ -1,1 }	{ 1,0,-1 }	{ 1,-1 }
<i>having_Sub_Domain</i>	<i>SSLfinal_State</i>	<i>Domain</i>
{ -1,0,1 }	{ -1,1,0 }	<i>registration</i>
		<i>length</i>
		{ -1,1 }
<i>Request_URL</i>	<i>URL_of_Anchor</i>	<i>Links_in_tags</i>
{ 1,-1 }	{ -1,0,1 }	{ 1,-1,0 }
<i>Redirect</i>	<i>on_mouseover</i>	<i>RightClick</i>
{ 0,1 }	{ 1,-1 }	{ 1,-1 }
<i>DNSRecord</i>	<i>web_traffic</i>	<i>Page_Rank</i>
{ -1,1 }	{ -1,0,1 }	{ -1,1 }
<i>Result</i>	<i>ADDRESS BAR</i>	<i>ABNORMAL</i>
{ -1,1 }	<i>BASED</i>	<i>BASED</i>
	<i>FEATURES</i>	<i>FEATURES</i>
<i>having_At_Symbol</i>	<i>double_slash_redirecting</i>	<i>Prefix_Suffix</i>
{ 1,-1 }	{ -1,1 }	{ -1,1 }
<i>Favicon</i>	<i>port</i>	<i>HTTPS_token</i>
{ 1,-1 }	{ 1,-1 }	{ -1,1 }
<i>SFH</i>	<i>Submitting_to_email</i>	<i>Abnormal_URL</i>
{ -1,1,0 }	{ -1,1 }	{ -1,1 }
<i>popUpWidnow</i>	<i>Iframe</i>	<i>age_of_domain</i>
{ 1,-1 }	{ 1,-1 }	{ -1,1 }
<i>Google_Index</i>	<i>Links_pointing_to_page</i>	<i>Statistical_report</i>
{ 1,-1 }	{ 1,0,-1 }	{ -1,1 }
<i>HTML AND</i>	<i>DOMAIN BASED</i>	CLASS
<i>JAVASCRIPT</i>	<i>FEATURES</i>	{ -1,1 }
<i>BASED</i>		
<i>FEATURES</i>		

Tabla 22: Valores posibles de cada campo caso C

Figura 33: Composición de los campos caso C.

APÉNDICE: Contribuciones

Como resultado de esta investigación se han presentado y publicado seis trabajos originales en conferencias internacionales y tres en revistas científicas de alto impacto. A continuación se describen brevemente cada uno de los artículos con sus respectivas contribuciones.

- **Use of a Semantic Language to Reduce the Indeterminacy in Agents Communication (2014).**

HORACIO PAGGI, COCHEZ, MICHAEL. 2014 International Conference on Mathematics and Computers in Sciences and in Industry (MCSI), Varna, Bulgaria. ISSN/ISBN: 978-1-4799-474. DOI: 10.1109/MCSI.2014.64. Publicado por IEEE.

Categoría: No categorizada dentro del portal CORE.

Aporte personal: El doctorando muestra cómo modelar el aprendizaje por refuerzo, dar información de contexto, etc. usando S-APL y así manejar la vaguedad e incertidumbre, es decir, la reducción del impacto de ellas se lograba por el uso de un lenguaje común. Cochez, como co-autor, era quien había desarrollado el lenguaje S-APL como parte de su tesis de maestría y colaboró brindando información que no estaba documentada sobre el mismo. Este trabajo permitió asimismo tomar contacto e investigar en forma conjunta con un académico totalmente desconocido, sólo a través de internet.

Contenido del trabajo: El objetivo fue explorar una herramienta, el lenguaje Semantic Agent Programming Language (S-APL) para la comunicación entre agentes que contuviera vaguedad y con resultados inciertos. Se muestra cómo usando dicho lenguaje, que posibilita incluir información de contexto en el mensaje así como realizar un aprendizaje por refuerzo básico, se puede reducir el impacto de la vaguedad e incertidumbres presentes en las comunicaciones entre agentes de software. o Contribución principal a la Tesis: Trabajo exploratorio. Se buscaba minimizar el impacto de la vaguedad e in-

certidumbre haciendo que los agentes usasen ese lenguaje de comunicación, que usaran información de contexto en sus mensajes y que aprendieran por refuerzo.

- **On the use of holonic agents in the design of information fusion systems (2014).**

HORACIO PAGGI, ÉLOI BOSSÉ, MIHAI CRISTIAN FLOREA, BASEL SOLAIMAN, FUSION 2014.

Categoría: CORE2014 , Fuente: ERA(2010) Rank: C del 2014 a la fecha, Primary Field Of Research: 0802 - Computation Theory and Mathematics

Aporte personal: El aporte del doctorando consistió en darle un enfoque realmente holónico a un modelo de fusión de información propuesto por Bossé et al.. Bossé y los demás co-autores aportaron conocimiento sobre distintos aspectos de fusión de información y modelos asociados. Este trabajo permitió al doctorando conocer el campo de la fusión de información y tomar contacto con un experto en la materia como es Bossé.

Contenido del trabajo: Se describe una interpretación del modelo JDL de fusión de información considerando que los componentes a distinto nivel pueden auto-organizarse libremente de forma que se generan holones.

Contribución principal a la Tesis: La fusión es propuesta como forma de tomar en cuenta la vaguedad, incertidumbre y otras imperfecciones de la información. Se observa que los holones van “mejorando” la información con cada nivel de procesamiento.

- **Cohesion and complex systems structure (2010).**

HORACIO PAGGI , FERNANDO ALONSO AMO. CODS 2010 - The 4th International Conference on Complex Distributed Systems (CODS 2010). Chongqing/China

Categoría: No categorizada dentro del portal CORE.

Aporte personal: Autor del trabajo.

Contenido del trabajo: Se analiza la formación de los holones considerando que cada agente es racional, que maneja la vaguedad e incertidumbre usando la teoría de probabilidades *fuzzy* (*fuzzy probabilities theory*) y que la toma de decisiones es hecha en base a la teoría de toma de decisiones *fuzzy* bajo riesgo (*fuzzy decision making under*

risk theory). A partir de allí se deducen las condiciones necesarias y suficientes para que un agente se una a un holón, considerando la utilidad o beneficio de la decisión. Asimismo se introduce el concepto de cohesión, definido como un cociente, que refleja lo que en otras áreas, por ejemplo psicología social, se conoce como cohesión de un grupo (una especie de aglutinante que lleva a las personas a pertenecer a un cierto grupo y que resiste a la separación de él). Ese valor definido de cohesión puede ser usado como una generalización de la confiabilidad del sistema. o Contribución evistprincipal a la Tesis: Trabajo exploratorio, se buscaba conocer en qué circunstancias se forman los holones (y cuando desaparecen).

- **Beliefs, Certainty and Complex Systems Structure (2010).**

HORACIO PAGGI , FERNANDO ALONSO AMO. The 2nd International Conference on Computer and Automation Engineering (ICCAE2010). Ciudad: Bali Island, Indonesia.

Categoría: No categorizada dentro del portal CORE.

Aporte personal: Autor

Contenido del trabajo: Se estudia las condiciones necesarias y suficientes para que se forme un holón a partir de otros en base a las creencias (*beliefs*) que tienen los holones de alcanzar cierta meta, considerando decisiones racionales e información vaga e incierta.

Contribución principal a la Tesis: Trabajo exploratorio, se buscaba conocer en qué circunstancias se forman los holones (y cuándo desaparecen) sin considerar cómo se modela la vaguedad e incertidumbre (es decir, sin utilizar teoría de la probabilidad, de conjuntos *fuzzy*, etc.).

- **Uncertainty and Randomness: A Holonic Approach (2010)**

HORACIO PAGGI, FERNANDO ALONSO AMO. The 2nd International Conference on Computer and Automation Engineering (ICCAE2010). Bali, Indonesia.

Categoría: No categorizada dentro del portal CORE.

Aporte personal: Autor.

Contenido del trabajo: Debido a la revisión que el autor realizó con respecto a los formalismos de representar vaguedad e incertidumbre, se bosqueja la idea de si la

separación que se suele hacer entre vaguedad e incertidumbre es sólo válida para entes en el mismo plano conceptual, es decir, si un holón formado por sub-holones que manejan información vaga no termina comportándose aleatoriamente, y, complementariamente, si de información con aleatoriedad en un nivel no se obtienen comportamientos vagos en un nivel superior.

Contribución principal a la Tesis: Trabajo exploratorio, se buscaba tratar de unificar las teorías de manejo de vaguedad e incertidumbre, quedando la pregunta planteada de si existe una transformación, expresable como una fórmula cerrada, que permita pasar de una a la otra al subir de nivel en la holarquía.

- **A Holonic Model of System for the Resolution of Incidents in the Software Engineering Projects (2009).**

HORACIO PAGGI, FERNANDO ALONSO. International Conference on Computer and Automation Engineering ICCAE '09. ISBN: 978-0-7695-3569-2 doi 10.1109/ICCAE.2009.16

Categoría: No categorizada dentro del portal CORE

Aporte personal: Autor

Contenido del trabajo: En este trabajo se presenta una aplicación concreta del concepto de holón, aplicada a una mesa de ayuda (*help desk*). Se esbozan las ideas de una organización que resuelve problemas en forma holónica. Debido a la alta especialización técnica de los componentes de una mesa de ayuda, su alta rotación y a la necesidad de automatizar dicha resolución, se plantea un modelo holónico donde los holones de menor nivel están formados por personas con sus herramientas de *hardware*, *software* y conocimientos, que se agrupan para resolver problemas.

Contribución principal a la Tesis: La idea es continuada en esta Tesis al ser la organización que resuelve problemas (o contesta consultas) una red de *peers*.

- **Indeterminacy Reduction in Agent Communication Using a Semantic Language (2015).**

HORACIO PAGGI , MICHAEL COCHEZ, WSEAS Transactions on Systems, 2015.
ISSN: 11092777

Categoría: Q4 in Computer Science Applications, Scimago SJR 2017 0.15

Aporte personal: Autor, se profundizan y completan los conceptos del trabajo presentado en la conferencia 2014 International Conference on Mathematics and Computers in Sciences and in Industry (MCSI). Cochez colaboró mejorando el estilo y ayudando a clarificar ideas.

Contenido del trabajo: Se amplía el trabajo presentado en la conferencia MCSI, agregándole como ejemplo el de una mesa de ayuda (help desk) donde la especificación de las incidencias y su solución están especificadas en S-APL.

Contribución principal a la Tesis: Trabajo exploratorio, que analiza cómo razonar, representar conocimiento y comunicarse usando un lenguaje semántico que permite manejar vaguedad e incertidumbre.

- **A multi-agent system for minimizing information indeterminacy within information fusion scenarios in peer-to-peer networks with limited resources (2018)**

Horacio Paggi, Javier Soriano, Juan Alfonso Lara:. Information Sciences. 451-452: 271-294 (2018)

Categoría: (JCR IF 4.832, Q1 in Computer Science, Information Systems). DOI: <https://doi.org/10.1016/j.ins.2018.04.019>

Aporte personal: Autor

Contenido del trabajo: Se plantea un modelo de sistema formado por pares (*peers*) los que realizan fusión de información con el fin de mejorar la calidad de la información que manejan. o Contribución principal a la Tesis: Define el modelo de comunicaciones y proceso de información que es central en esta Tesis.

- **Structures generated in a multiagent system performing information fusion in peer-to-peer resource-constrained networks (2018).**

HORACIO PAGGI, JUAN ALFONSO LARA y JAVIER SORIANO. J. Neural Computing and Applications (2018).

Categoría: (JCR IF 4,213, Q1 in Computer Science, Artificial Intelligence). DOI: <https://doi.org/10.1007/s00521-018-3818->

Aporte personal: Autor

Contenido del trabajo: Se estudia la emergencia de estructuras holónicas en una red *peer-to-peer* cuando se mejora la calidad de información circulante usando fusión de información. o Contribución principal a la Tesis: Es un trabajo central para esta Tesis.