



**ACCIÓN ESTRATÉGICA EN ECONOMÍA Y SOCIEDAD DIGITAL 2013
“INDUSTRIAS DEL FUTURO. INFRAESTRUCTURAS DE INTERNET DEL
FUTURO”**

BIG MARKET

**Plataforma Big Data para simular y evaluar técnicas de
marketing en entornos realistas**

P6.2 Iteración inicial de Prototipos y Evaluación

Paradigma Tecnológico (beneficiario)
Universidad Politécnica de Madrid (subcontratada)

Tabla de contenidos

Índice de ilustraciones.....	3
Resumen ejecutivo.....	4
Información del Documento.....	5
1. Extensión de BigTweet.....	6
2. Javadoc de BigTweet.....	7
3. Experimentos de evaluación.....	9
4. Resumen final y conclusión.....	12
5. Referencias.....	13

Índice de ilustraciones

Ilustración 1: Diagrama de Clases UML de la interfaz gráfica de BigTweet extendida para la evaluación.....	6
Ilustración 2: Resultados experimentales. Modelo baseline M1 vs modelo propuesto M2 usando los datasets de Obama y Palin.....	11

Resumen ejecutivo

Este entregable describe una primera iteración de los prototipos, integrando el resultado del resto de paquetes de trabajo técnicos, y se evalúa con el los datos establecidos en E6.1.

Información del Documento

Proyecto FIT Número	TSI-100102-2013-80	Acrónimo	Big Market
Título completo	Plataforma Big Data para simular y evaluar técnicas de marketing en entornos realistas		
URL	http://www.paradigmatecnologico.com/portfolio/bigmarket/		
URL del documento			

Entregable	Número	E6.2	Título	Iteración inicial de Prototipos y Evaluación
Paquete de Trabajo	Número	PT6	Título	Simulación social basada en Big Data
Tarea	Número	T6.2	Título	Iteración inicial de Prototipos y Evaluación

Fecha de Entrega	Contractual	27/03/2015 (M15)	Entregado	
Estado	Versión final 27/03/2015 (M15)		Final	
Tipo	Informe	Demo	X Otro	
Nivel de Diseminación	Público	X Consorcio		
Resumen (para diseminación)	Primera iteración de los prototipo y evaluación.			
Palabras Clave	Justificación			

Autores (Partner)	UPM			
Responsable de Autoría	Partner	UPM	Email	cif@gsi.dit.upm.es

1. Extensión de BigTweet

El prototipo BigTweet presentado en P5.2 se ha ampliado con estas estrategias. Para este propósito, se han incluido varias clases para la evaluación como se muestra en el diagrama UML de la figura 1: BTSimBatch, BTSimBatchSB, BTSimBatchCWRD, BatchExperimentResults, StudyingBeacons, ComparingWithRealData, MonitorAgent y Evaluador. Además, el lenguaje de programación R se ha empleado para realizar operaciones estadísticas avanzadas.

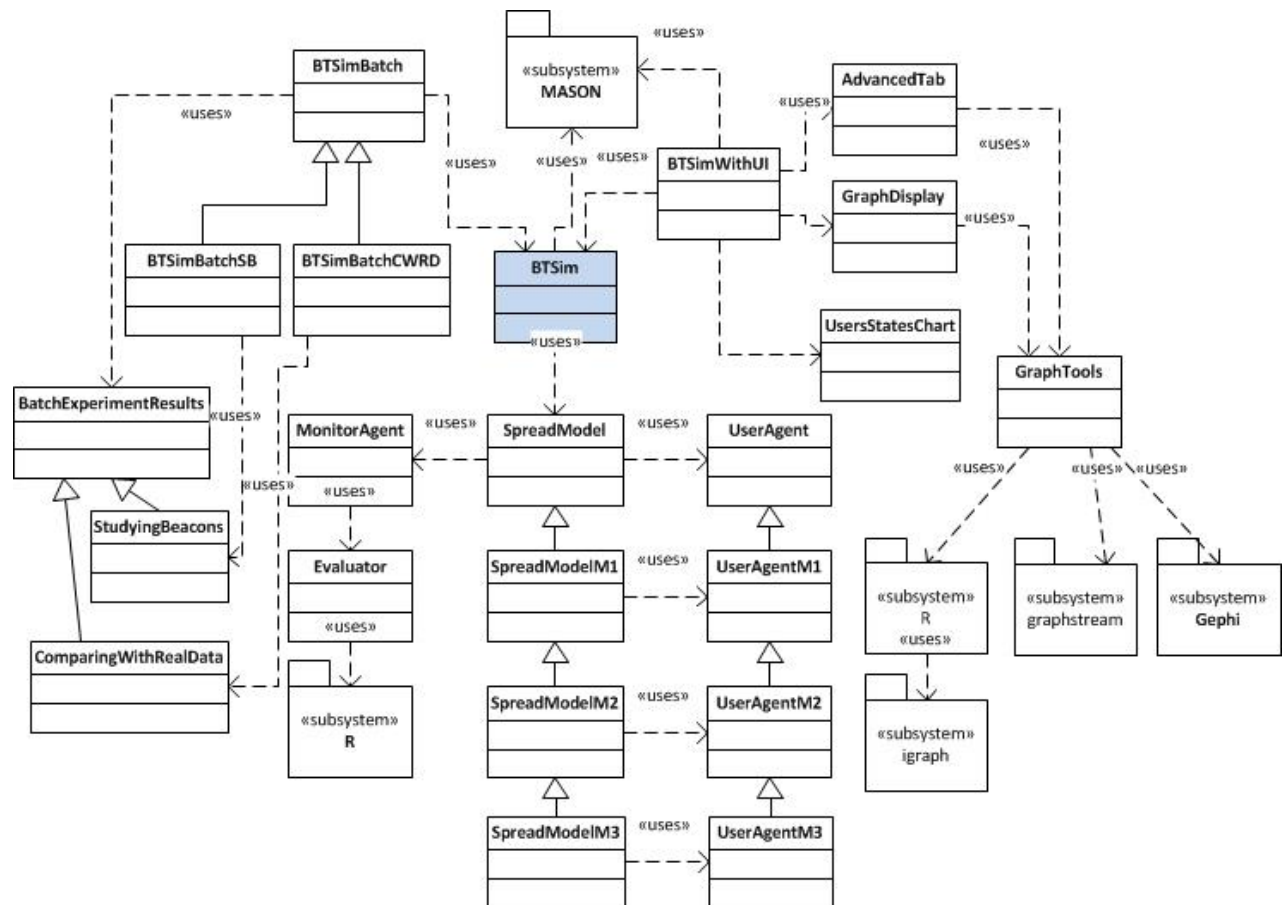


Ilustración 1: Diagrama de Clases UML de la interfaz gráfica de BigTweet extendida para la evaluación

2. Javadoc de BigTweet

La documentación de referencia de BigTweet en formato javadoc, que incluye el código de los experimentos está disponible online en: <http://www.gsi.dit.upm.es/~eserrano/BigMarket/P6.2>.

Incluye las siguientes clases java:

- [AdvancedTab](#)
- [BatchExperimentsResults](#)
- [BTSim](#)
- [BTSimBatch](#)
- [BTSimBatchStudyingBeacons](#)
- [BTSimComparingToRealData](#)
- [BTSimWithUI](#)
- [Clicks](#)
- [ComparingToRealData](#)
- [EvaluationTools](#)
- [GraphDisplay](#)
- [GraphTools](#)
- [MonitorAgent](#)
- [SpreadModel](#)
- [SpreadModelM1](#)
- [SpreadModelM1.State](#)
- [SpreadModelM2](#)
- [SpreadModelM3](#)
- [StudyingBeacons](#)
- [UserAgent](#)

- [UserAgentM1](#)
- [UserAgentM2](#)
- [UserAgentM3](#)
- [UsersStatesChart](#)

3. Experimentos de evaluación

Como se explicó en E5.1, la validación evalúa el realismo del modelo construido a partir de los datos obtenidos del sistema objetivo. Para ello, se comparan el modelo de propagación de habladurías y el modelo propuesto (ver P5.3) utilizando los dos conjuntos de datos de Twitter explicados en T6.1.

Estos experimentos comparan el número de usuarios que avalan y niegan un rumor en la simulación con el número de estos usuarios en los datos reales. Por lo tanto, la siguiente métrica de distancia se utiliza para validar el realismo de las simulaciones:

$$d(endorsers, simulation, dataset) + d(deniers, simulation, dataset)$$

donde d calcula la distancia euclídea entre el número de usuarios (nu) de un tipo específico (*endorser* o *denier*) en la simulación y el conjunto de datos (Obama o Palin) para los días considerados ($nDays$):

$$\left[\sum_{day=0}^{nDays} (nu(type, simulation, day) - nu(type, dataset, day))^2 \right]^{\frac{1}{2}}$$

El número de endosantes y usuarios denier se calcula de manera diferente para los conjuntos de datos, el modelo de referencia y el modelo propuesto. En los conjuntos de datos, un usuario se cuenta como endosante o denier si su último tweet fue etiquetado como endorsers o denies, respectivamente. En el enfoque basal, los usuarios infectados cuentan como endosantes y, vacunados y curados como negadores. En el modelo propuesto, los agentes curados se cuentan como usuarios respaldando el rumor junto con los agentes infectados; Y sólo los agentes vacunados se cuentan como usuarios que niegan el rumor.

Parameter	Min	Max	Inc
probInfect	0.01	0.1	0.005
probAcceptDen y	0.01	0.1	0.005

timeLag ()	0	23	1
probMakeDenier ()	0.01	0.1	0.005
random seed	1	20	1
Parameter	Fixde value		
Users	1000		
InitiallyInfected	2		
MaxLinkPerNode	10		

Tabla 1. Valores de los parámetros de los modelos. M1: sólo baseline; M2: sólo el modelo propuesto.

La Tabla 1 muestra los parámetros empleados para los dos modelos de difusión explicados en P5.2. Algunos parámetros se especifican con mínimo, máximo e incremento, mientras que otros tienen un valor fijo. Las combinaciones de los parámetros han sido cerca de 173K experimentos para la línea base y cerca de 170k experimentos para el modelo propuesto.

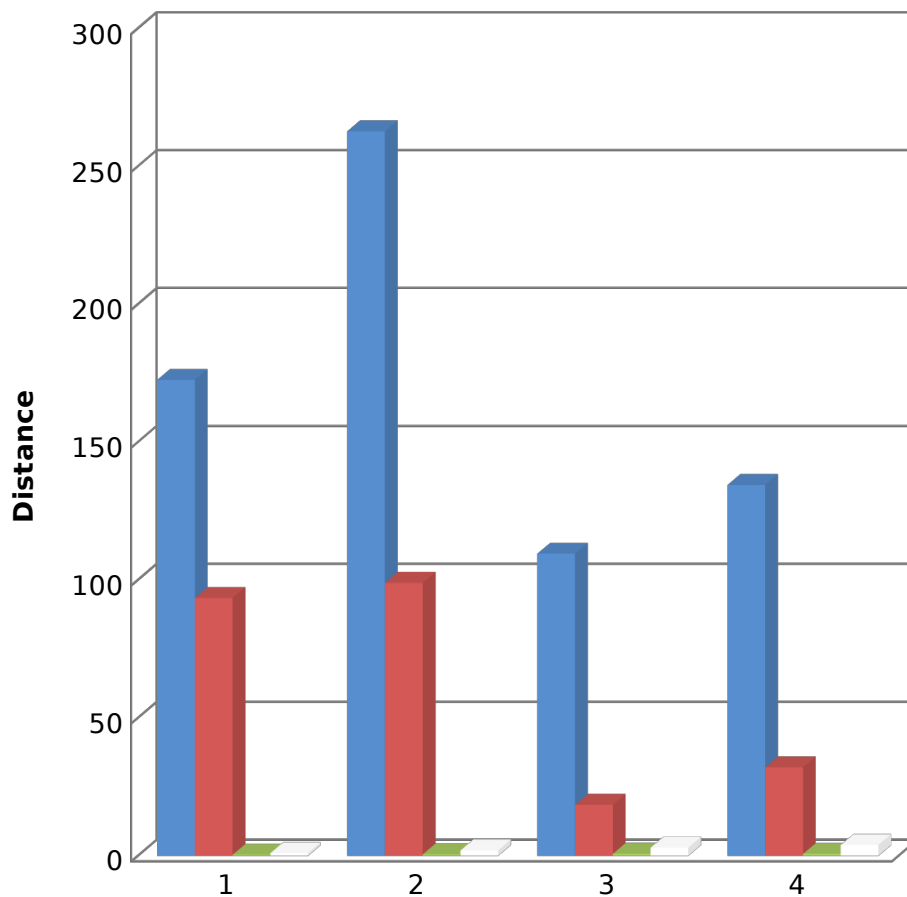


Ilustración 2: Resultados experimentales. Modelo baseline M1 vs modelo propuesto M2 usando los datasets de Obama y Palin

La Figura 2 muestra los principales resultados. La figura muestra las distancias para los datos de Obama y Palin: (1) en el mejor de los casos alcanzados por los modelos; (2) y, en la mejor media de las distancias considerando todas las semillas al azar para un conjunto de valores de los parámetros¹. Estos resultados muestran que el modelo propuesto alcanza reducciones de la distancia entre el 45,80%, el mejor caso de Obama y el 83,07%, el mejor caso de Palin. Estos resultados experimentales apoyan la hipótesis de que es más realista considerar que los usuarios que han difundido un rumor no difundirán anti-rumores en Twitter como lo hace el nuevo modelo introducido en P5.2. Los exámenes de experimentos y los resultados de experimentos extendidos están disponibles en línea en la web de material adicional [Serrano et al., B].

4. Resumen final y conclusión

Esta evaluación desafía los modelos de propagación de rumores clásicos asumiendo que los usuarios que se dan cuenta de que han difundido un rumor falso en Twitter típicamente: (1) no difundirán rumores anti-rumores, o (2) no habrá evidencia empírica de la retracción. Por lo tanto, los usuarios recuperados no afectarán la recuperación de sus vecinos. El análisis exploratorio de datos de dos conjuntos de datos de rumores de Twitter sobre Obama y Palin explicado en E6.1 apoya esta hipótesis. Concretamente, la mayoría de la información del usuario proviene de un solo tweet que permite a los investigadores considerarlos rumores endosantes o negadores, haciendo que la percepción de "recuperación" no esté disponible. El nuevo modelo de propagación de rumores basado en agentes se compara con uno de los primeros modelos de propagación de información errónea basados en agentes de simulación social (ABSS) para el caso de Twitter. Los resultados experimentales muestran que el nuevo modelo es capaz de reducir entre el 45% y el 83% la distancia con los dos conjuntos de datos de Twitter estudiados.

5. Referencias

- [wik, a] Agent-based modeling software frameworks list.. Accessed February 1, 2015.
http://en.wikipedia.org/wiki/ABM_Software_Comparison
- [wik, b] Social network analysis software.. Accessed February 1, 2015.
http://en.wikipedia.org/wiki/Social_network_analysis_software
- [twi, a] Twitter REST API console website. Accessed February 1, 2015.
<https://dev.twitter.com/rest/tools/console>.
- [twi, b] Twitter REST API documentation website. Accessed February 1, 2015.
<https://dev.twitter.com/overview/documentation>.
- [twi, c] Twitter terms of use website. Accessed February 1, 2015.
<https://dev.twitter.com/overview/terms/policy>.
- [Abraham et al., 2010] Abraham, A., Hassanien, A.-E., and Snasel, V. (2010). Computational social network analysis. *Computational Social Network Analysis, Computer Communications and Networks*, 1.
- [Andrienko and Andrienko, 2005] Andrienko, N. and Andrienko, G. (2005). *Exploratory Analysis of Spatial and Temporal Data: A Systematic Approach*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [Beck et al., 2001] Beck, K., Beedle, M., van Bennekum, A., Cockburn, A., Cunningham, W., Fowler, M., Grenning, J., Highsmith, J., Hunt, A., Jeffries, R., Kern, J., Marick, B., Martin, R. C., Mellor, S., Schwaber, K., Sutherland, J., and Thomas, D. (2001). *Manifesto for Agile Software Development*. Online version: <http://agilemanifesto.org/>; letzter Zugriff 08.07.2010).
- [Bothos et al., 2010] Bothos, E., Apostolou, D., and Mentzas, G. (2010). Using social media to predict future events with agent-based markets. *IEEE Intelligent Systems*, 25(6):50–58.
- [Buchanan, 2009] Buchanan, M. (2009). Economics: Meltdown modelling. *Nature*, 460(7256):680.

[Campuzano et al., 2015] Campuzano, F., Garcia-Valverde, T., Botia, J. A., and Serrano, E. (2015). Generation of human computational models with machine learning. *Information Sciences*, 293(0):97 – 114.

[Campuzano et al., 2014] Campuzano, F., Garcia-Valverde, T., Serrano, E., and Botía, J. A. (2014). Generation of human computational models with knowledge engineering. *Engineering Applications of Artificial Intelligence*, 35(0):259 – 276.

[Cha et al., 2010] Cha, M., Haddadi, H., Benevenuto, F., and Gummadi, K. (2010). Measuring user influence in twitter: The million follower fallacy. In *4th International AAAI Conference on Weblogs and Social Media (ICWSM)*.

[Challenger et al., 2014] Challenger, M., Demirkol, S., Getir, S., Mernik, M., Kardas, G., and Kosar, T. (2014). On the use of a domain-specific modeling language in the development of multiagent systems. *Engineering Applications of Artificial Intelligence*, 28(0):111 – 141.

[Cohen,] Cohen, H. 72 marketing definitions. . Accessed February 1, 2015. <http://heidicohen.com/marketing-definition/>

[De Domenico et al., 2013] De Domenico, M., Lima, A., Mougél, P., and Musolesi, M. (2013). The Anatomy of a Scientific Rumor. *Scientific Reports*, 3.

[Farmer and Foley, 2009] Farmer, J. D. and Foley, D. (2009). The economy needs agent-based modelling. *Nature*, 460(7256):685–686.

[Flentge et al.,] Flentge, F., Polani, D., and Uthmann, T. Modelling the emergence of possession norms using memes. *J. Artificial Societies and Social Simulation*, (4).

[Frias-Martinez and Frias-Martinez, 2014] Frias-Martinez, V. and Frias-Martinez, E. (2014). Spectral clustering for sensing urban land use using twitter activity. *Engineering Applications of Artificial Intelligence*, 35(0):237 – 245.

[Garcia-Valverde et al., 2012] Garcia-Valverde, T., Campuzano, F., Serrano, E., Villa, A., and Botia, J. A. (2012). Simulation of human behaviours for the validation of ambient intelligence services: A methodological approach. *Journal of Ambient Intelligence and Smart Environments*, 4(3):163–181.

- [Gatti et al., 2013] Gatti, M. A. d. C., Appel, A. P., dos Santos, C. N., Pinhanez, C. S., Cavalin, P. R., and Neto, S. B. (2013). A simulation-based approach to analyze the information diffusion in microblogging online social network. In *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World*, WSC '13, pages 1685–1696, Piscataway, NJ, USA. IEEE Press.
- [Gilbert and Troitzsch, 2005] Gilbert, N. and Troitzsch, K. G. (2005). *Simulation for the Social Scientist*. Open University Press.
- [Gupta et al., 2013a] Gupta, A., Lamba, H., and Kumaraguru, P. (2013a). \$1.00 per RT #BostonMarathon #PrayForBoston: Analyzing fake content on twitter. San Francisco, CA.
- [Gupta et al., 2013b] Gupta, A., Lamba, H., Kumaraguru, P., and Joshi, A. (2013b). Faking sandy: Characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22Nd International Conference on World Wide Web Companion*, WWW '13 Companion, pages 729–736, Republic and Canton of Geneva, Switzerland. International World Wide Web Conferences Steering Committee.
- [Gutiérrez et al., 2013] Gutiérrez, C., García-Magariño, I., Serrano, E., and Botía, J. A. (2013). Robust design of multi-agent system interactions: A testing approach based on pattern matching. *Engineering Applications of Artificial Intelligence*, 26(9):2093 – 2104.
- [Hethcote, 2000] Hethcote, H. W. (2000). The mathematics of infectious diseases. *SIAM Review*, 42:599–653.
- [Holubiec et al., 2012] Holubiec, J., Szkatula, G., and Wagner, D. (2012). A knowledge-based model of parliamentary election. *Information Sciences*, 202:24–40.
- [Jin et al., 2013] Jin, F., Dougherty, E., Saraf, P., Cao, Y., and Ramakrishnan, N. (2013). Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, SNAKDD '13, pages 8:1–8:9, New York, NY, USA. ACM.
- [Khamis and Nagi, 2013] Khamis, M. A. and Nagi, K. (2013). Designing multi-agent unit tests using systematic test design patterns-(extended version). *Engineering Applications of Artificial Intelligence*, 26(9):2128 – 2142.

[Kilikki et al., 2014] Kilikki, O., Kangasrääsio, A., Nikkilä, R., Alahäivälä, A., and Seilonen, I. (2014). Agent-based modeling and simulation of a smart grid: A case study of communication effects on frequency control. *Engineering Applications of Artificial Intelligence*, 33(0):91 – 98.

[Kitchenham et al., 2009] Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., and Linkman, S. (2009). Systematic literature reviews in software engineering - a systematic literature review. *Inf. Softw. Technol.*, 51(1):7–15.

[Kostka et al., 2008] Kostka, J., Oswald, Y. A., and Wattenhofer, R. (2008). Word of Mouth: Rumor Dissemination in Social Networks Structural Information and Communication Complexity. In Shvartsman, A. A. and Felber, P., editors, *Structural Information and Communication Complexity*, volume 5058 of *Lecture Notes in Computer Science*, chapter 16, pages 185–196. Springer Berlin / Heidelberg, Berlin, Heidelberg.

[Kwon et al., 2013a] Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y. (2013a). Aspects of rumor spreading on a microblog network. In Jatowt, A., Lim, E.-P., Ding, Y., Miura, A., Tezuka, T., Dias, G., Tanaka, K., Flanagan, A., and Dai, B., editors, *Social Informatics*, volume 8238 of *Lecture Notes in Computer Science*, pages 299–308. Springer International Publishing.

[Kwon et al., 2013b] Kwon, S., Cha, M., Jung, K., Chen, W., and Wang, Y. (2013b). Prominent features of rumor propagation in online social media. In Xiong, H., Karypis, G., Thuraisingham, B. M., Cook, D. J., and Wu, X., editors, *2013 IEEE 13th International Conference on Data Mining, Dallas, TX, USA, December 7-10, 2013*, pages 1103–1108. IEEE Computer Society.

[Landwehr and Carley, 2014] Landwehr, P. and Carley, K. (2014). Social media in disaster relief. In Chu, W. W., editor, *Data Mining and Knowledge Discovery for Big Data*, volume 1 of *Studies in Big Data*, pages 225–257. Springer Berlin Heidelberg.

[Li et al., 2008] Li, X., Mao, W., Zeng, D., and Wang, F.-Y. (2008). Agent-based social simulation and modeling in social computing. In Yang, C., Chen, H., Chau, M., Chang, K., Lang, S.-D., Chen, P., Hsieh, R., Zeng, D., Wang, F.-Y., Carley, K., Mao, W., and Zhan, J., editors, *Intelligence and Security Informatics*, volume 5075 of *Lecture Notes in Computer Science*, pages 401–412. Springer Berlin Heidelberg.

[Liu and Chen, 2011] Liu, D. and Chen, X. (2011). Rumor propagation in online social networks like twitter – a simulation study. In *Proceedings of the 2011 Third International Conference on*

Multimedia Information Networking and Security, MINES '11, pages 278–282, Washington, DC, USA. IEEE Computer Society.

[Long and Wong, 2014] Long, C. and Wong, R. C.-W. (2014). Viral marketing for dedicated customers. *Information Systems*, 46(0):1 – 23.

[Luke et al., 2004] Luke, S., Cioffi-Revilla, C., Panait, L., and Sullivan, K. (2004). Mason: A new multi-agent simulation toolkit. In *Proceedings of the 2004 Swarmfest Workshop*.

[Mavridis and Symeonidis, 2014] Mavridis, T. and Symeonidis, A. L. (2014). Semantic analysis of web documents for the generation of optimal content. *Engineering Applications of Artificial Intelligence*, 35(0):114 – 130.

[Mendoza et al., 2010] Mendoza, M., Poblete, B., and Castillo, C. (2010). Twitter under crisis: Can we trust what we rt? In *Proceedings of the First Workshop on Social Media Analytics*, SOMA '10, pages 71–79, New York, NY, USA. ACM.

[Nekovee et al., 2007] Nekovee, M., Moreno, Y., Bianconi, G., and Marsili, M. (2007). Theory of rumour spreading in complex social networks. *Physica A: Statistical Mechanics and its Applications*, 374(1):457 – 470.

[North et al., 2006] North, M. J., Collier, N. T., and Vos, J. R. (2006). Experiences creating three implementations of the repast agent modeling toolkit. *ACM Trans. Model. Comput. Simul.*, 16(1):1–25.

[Nwana, 1996] Nwana, H. S. (1996). Software agents: An overview. *Knowledge Engineering Review*, 11:205–244.

[Placzek, 2014] Placzek, B. (2014). A self-organizing system for urban traffic control based on predictive interval microscopic model. *Engineering Applications of Artificial Intelligence*, 34(0):75 – 84.

[Qazvinian et al., 2011] Qazvinian, V., Rosengren, E., Radev, D. R., and Mei, Q. (2011). Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, EMNLP '11, pages 1589–1599, Stroudsburg, PA, USA. Association for Computational Linguistics.

- [Rand and Rust, 2011] Rand, W. and Rust, R. T. (2011). Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing*, 28(3):181–193.
- [Rolla and Curado, 2013] Rolla, V. G. and Curado, M. (2013). A reinforcement learning-based routing for delay tolerant networks. *Engineering Applications of Artificial Intelligence*, 26(10):2243 – 2250.
- [Rossant, 2013] Rossant, C. (2013). *Learning IPython for Interactive Computing and Data Visualization*. Packt Publishing.
- [Russell, 2011a] Russell, M. A. (2011a). *21 Recipes for Mining Twitter*. O’Reilly Media, 1 edition.
- [Russell, 2011b] Russell, M. A. (2011b). *Mining the Social Web*. O’Reilly Media, 1 edition.
- [Seo et al., 2012] Seo, E., Mohapatra, P., and Abdelzaher, T. (2012). Identifying rumors and their sources in social networks.
- [Serrano and Botia, 2013] Serrano, E. and Botia, J. (2013). Validating ambient intelligence based ubiquitous computing systems by means of artificial societies. *Information Sciences*, 222(0):3 – 24.
- [Serrano et al., a] Serrano, E., Iglesias, C. A., and Garijo, M. BigTweet website, simulator for rumor spreading models and rumor control strategies in Twitter with support for Big Data technologies. Accessed February 1, 2015. <https://github.com/gsi-upm/BigTweet>.
- [Serrano et al., b] Serrano, E., Iglesias, C. A., and Garijo, M. Paper additional material: datasets, simulations code, and extended experiments. . Accessed February 1, 2015. <http://www.gsi.dit.upm.es/~eserrano/BigMarket/EAAI2015/>
- [Serrano et al., 2014a] Serrano, E., Moncada, P., Garijo, M., and Iglesias, C. A. (2014a). Evaluating social choice techniques into intelligent environments by agent based social simulation. *Information Sciences*, 286(0):102–124.
- [Serrano et al., 2014b] Serrano, E., Poveda, G., and Garijo, M. (2014b). Towards a holistic framework for the evaluation of emergency plans in indoor environments. *Sensors*, 14(3):4513–4535.

[Serrano et al., 2013] Serrano, E., Rovatsos, M., and Botía, J. A. (2013). Data mining agent conversations: A qualitative approach to multiagent systems analysis. *Information Sciences*, 230(0):132 – 146.

[Shah and Zaman, 2011] Shah, D. and Zaman, T. (2011). Rumors in a network: Who's the culprit? *IEEE Transactions on Information Theory*, 57(8):5163–5181.

[Shamshirband et al., 2013] Shamshirband, S., Anuar, N. B., Kiah, M. L. M., and Patel, A. (2013). An appraisal and design of a multi-agent system based cooperative wireless intrusion detection computational intelligence technique. *Engineering Applications of Artificial Intelligence*, 26(9):2105 – 2127.

[Starbird et al., 2014] Starbird, K., Maddock, J., Orand, M., Achterman, P., and Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on twitter after the 2013 boston marathon bombing. In *iConference 2014 Proceedings* (p. 654 - 662).

[Statell,] Statell, G. Big data and simulations are transforming marketing.. Accessed February 1, 2015. <http://www.businessinsider.com/big-data-in-marketing-2013-5>

[Tisue and Wilensky, 2004] Tisue, S. and Wilensky, U. (2004). NetLogo: A Simple Environment for Modeling Complexity.

[Tripathy et al., 2010] Tripathy, R. M., Bagchi, A., and Mehta, S. (2010). A study of rumor control strategies on social networks. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management, CIKM '10*, pages 1817–1820, New York, NY, USA. ACM.

[Valecha et al., 2013] Valecha, R., Oh, O., and Rao, H. R. (2013). An exploration of collaboration over time in collective crisis response during the haiti 2010 earthquake. In Baskerville, R. and Chau, M., editors, *Proceedings of the International Conference on Information Systems, ICIS 2013, Milano, Italy, December 15-18, 2013*. Association for Information Systems.

[Weng et al., 2013] Weng, L., Menczer, F., and Ahn, Y.-Y. (2013). Virality prediction and community structure in social networks. *Scientific Reports*, 3.

[Woolridge, 2001] Woolridge, M. (2001). *Introduction to Multiagent Systems*. John Wiley & Sons, Inc., New York, NY, USA.

[Yang et al.,] Yang, S. Y., Liu, A., and Mo, S. Y. K. Twitter financial community modeling using agent based simulation. SSRN scholarly paper, Rochester, NY. IEEE Computational Intelligence in Financial Engineering and Economics, London, 2013.

[Zhao et al., 2013] Zhao, L., Cui, H., Qiu, X., Wang, X., and Wang, J. (2013). {SIR} rumor spreading model in the new media age. *Physica A: Statistical Mechanics and its Applications*, 392(4):995 – 1003.