Network Science, Web Science and Internet Science: comparing interdisciplinary areas

Thanassis Tiropanis[1], Wendy Hall[1], Jon Crowcroft[2], Noshir Contractor[3], Leandros Tassiulas[4]

[1] University of Southampton
[2] University of Cambridge
[3] Northwestern University
[4] University of Thessaly

ABSTRACT
In this paper, we examine the fields of Network Science, Web Science and Internet Science. All three areas are interdisciplinary, and since the Web is based on the Internet, and both the Web and the Internet are networks, there is perhaps confusion about the relationship between them. We study the extent of overlap and ask whether one includes the others, or whether they are all part of the same larger domain. This paper provides an account of the emergence of each of these areas and outlines a framework for comparison. Based on this framework, we discuss these overlaps and propose directions for harmonization of research activities.

INTRODUCTION

The observation of patterns that characterise networks, from biological to technological and social, and the impact of the Web and the Internet on society and business have motivated interdisciplinary research to advance our understanding of these systems. Their study has been the subject of Network Science research for a number of years. However, more recently we have witnessed the emergence of two new interdisciplinary areas: Web Science and Internet Science.

Network Science

Network science can be traced to its mathematical origins dating back to Leonard Euler's seminal work on graph theory (Euler, 1741) in the 18th century and to its social scientific origins two centuries later by the psychiatrist Jacob Moreno's (1953) efforts to develop "sociometry." Soon thereafter, the mathematical framework offered by graph theory was also picked up by psychologists (Bavelas, 1950), anthropologists (Mitchell, 1956), and other social scientists to create an interdiscipline called Social Networks. The inter-discipline of Social Networks expanded even further towards the end of the 20th century with an explosion of interest in exploring networks in biological, physical, and technological systems. The term Network Science emerged as an interdisciplinary area that draws on disciplines such as physics, mathematics, computer science, biology, economics and sociology to encompass networks that were not necessarily social (Barabási, A.-L., & Albert, 1999; Newman, 2010; Watts, 2004). The study of networks involves developing explanatory models to understand the emergence of networks, building predictive models to anticipate the evolution of networks, and constructing prescriptive models to optimize the outcomes of networks. One of the main tenets of Network Science is to identify common underpinning principles and laws that apply across very different networks and explore why in some cases those patterns vary. The Internet and the Web, given their spectacular growth and impact, are networks that have captured the imagination of many Network scientists (Easley & Kleinberg, 2010). In addition, the emergence of online social networks and the potential to study online interactions on a massive, global scale hold the promise of further, potentially invaluable insights to network scientists on network evolution (Monge and Contractor, 2003).

Web Science

Web Science (Berners-Lee et al., 2006a) is an interdisciplinary area of much more recent vintage that studies the Web not only at the level of small technological innovations (micro level) but also as a phenomenon that affects societal and commercial activities globally (macro level); to a large extent, it can be considered the theory and practice of social machines on the Web. Social machines were conceptualised by Tim Berners-Lee in 1999 as artefacts where people do the creative work and machines intermediate (Berners-Lee, 1999). Semantic Web and linked data technologies can provide the means for knowledge representation and reasoning and enable further support for social machines (Hendler & Berners-Lee, 2009).

Studying the Web and its impact requires an interdisciplinary approach that focuses not only on the technological level but also on the societal, political and commercial levels. Establishing the relationship between these levels, understanding how they influence each other, investigating potential underpinning laws and exploring ways to leverage this relationship in different domains of human activity is a large part of the Web Science research agenda. Web Science draws on disciplines that include the social sciences, such as anthropology, communication, economics, law, philosophy, political science, psychology and sociology but also computer science and engineering. A major focus of the Web Science research agenda is to understand how the Web is evolving as a socio-technical phenomenon and how we can ensure that it will continue to evolve and benefit society in the years to come.

Internet Science

The Internet has provided the infrastructure on which much of human activity has become heavily dependent. After only a few decades of Internet development it is self-evident that, if the Internet became unavailable, the consequences for society, commerce, the economy, defence and government would be highly disruptive. The success of the Internet has often been attributed to its distributed governance model, the principle of network neutrality and its openness (Economides, 2008). At the same time, concerns related to privacy, security, openness and sustainability are raised and researched as they are often at the centre of contestations on the Internet (Clark et al., 2002). The Internet can be seen as an infrastructure, the social value of which needs to be safeguarded (Frischmann, 2012). It is the infrastructure that enabled the evolution of the Web along with P2P applications, more recently the Cloud, and, in the near future, the Internet of Things. It has been argued that the infrastructural layer of the Internet and that of the Web need to be kept separately to foster innovation (Berners-Lee, 2010). A recent study (Blackman et al., 2010) identified a number of principled directions along which the Internet needs to evolve; those include availability, inclusiveness, scalability, sustainability, openness, security, privacy and resilience. This motivates the need for multidisciplinary research on Internet Science that seeks to understand the psychological, sociological and economic implications of the Internet's evolution along these principled directions. Hence Internet Science is an emerging interdisciplinary area that brings together scientists in network engineering, computation, complexity, security, trust, mathematics, physics, sociology, economics, political sciences and law. This approach is very well exemplified by the early Internet Topology study (Faloutsos, 1999).

Interdisciplinary relationships

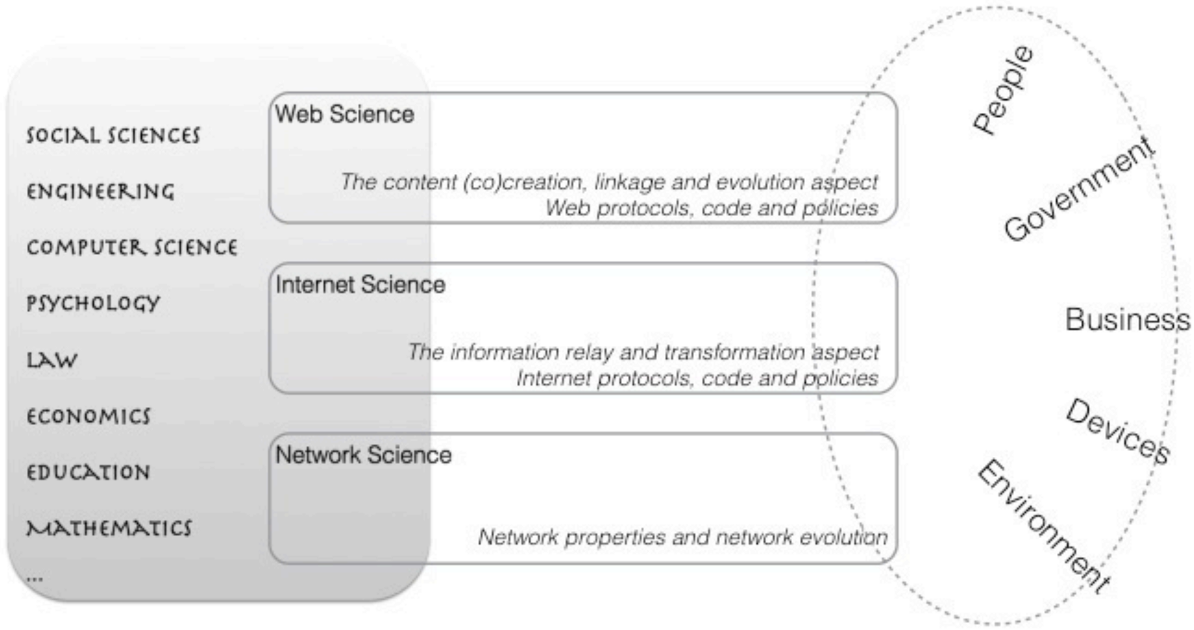

Figure 1: Web, Internet and Network Science aspects.

All three areas draw on a number of disciplines for the study, respectively, the nature and impact of the Web, the Internet and networks in general on government, business, people, devices and the environment. However, each of them examines how those actors co-create and evolve in distinct, unique ways as shown on Figure 1. For Web Science it is the aspect of linking those actors and the content with which they interact making associations between them and interpreting them. For Internet Science it is the aspect of communication among actors and resources as processes that can shape information relay and transformation. For Network Science, it is the aspect of how the above entities, when considered to be part of a network, exhibit certain characteristics and might adhere to underpinning laws that can help understand their evolution.

However, to understand better the similarities and differences between these areas and to establish the potential for synergies, a framework for a more detailed comparison is needed.

A COMPARISON OF INTERDISCIPLINARY AREAS

It takes only a quick read through a short description of each of these interdisciplinary areas (Berners-Lee et al., 2006b; The EINS Consortium, 2012; Watts, 2004) for one to realise that, to a very large extent, they all draw from very similar sets of disciplines. Venn diagrams that have been used to illustrate the involvement of different disciplines in each area are indicative of this overlap. For example, psychology and economics are considered relevant to Network Science (Steen, 2011), Internet Science (Blackman et al., 2010) and Web Science (Hendler & Berners-Lee, 2009). This can give rise to certain questions such as “if there is so much overlap, aren’t these areas one and the same?” or “would they all merge in the future?” Other questions include: “which community is more relevant to my research?” or “what developments could we expect from each area in the future?” To explore those questions we propose a framework of examining those interdisciplinary areas, which includes looking at the way that these communities have formed, and the different languages of discourse that these communities have employed in their research.

Community formation

Although not all three interdisciplinary domains were established at the same time, one can argue that research in those areas dates back before their official starting date. At the same time, one can also argue that there are differences in how communities around those domains emerged.

The formation of the Social Networks community can be traced back to a series of social network conferences that started in the 70s (Freeman, 2004) with an important conference in Dartmouth in 1975 which brought together sociologists, anthropologists, social psychologists, and mathematicians from the United States and Europe. This was followed by Lin Freeman’s launch of the Social Networks journal in 1978, and Barry Wellman founding INSNA (International Network for Social Network Analysis) in 1976 and its annual “Sunbelt Social Networks” conference in 1981. Beginning in the 1990s, the social scientists were joined by a large and growing influx of scholars from the physical and life sciences who began exploring networks in social systems. This effort was acknowledged and further catalysed by the launch of annual Network Science, NetSci, conference in 2006, a major infusion of funding in 2008 from the Army Research Laboratory for the development of an interdisciplinary Network Science Collaborative Technology Alliance (NS-CTA), and the launch of the Network Science journal in 2013. Clearly there was already a community in place, which engaged in interdisciplinary work long before those initiatives; one can argue that a hybrid bottom-up and top-down approach is the community formation model that was followed for Network Science.

For the Web Science community, it was around 2006 when it was realised that understanding the impact of the Web was essential to safeguard its development in the future. The Web Science Research Initiative (WSRI) was established in 2006 and later developed into the Web Science Trust (WST) as part of the top-down approach to the formation of the Web Science community. The WSRI raised a banner for those who were engaged in research on the Web as a socio-technical phenomenon, including the social network research community. A similar community formation model was followed for Internet Science, where the European Network of Excellence in Network Science (EINS)[1] is one of the most significant activities to bring together the research community in this area. Areas such as privacy and network neutrality have been highlighted as priorities in the Internet Science agenda.

It can be argued that the top-down model of community formation can accelerate research in emergent interdisciplinary areas but, in order to be successful, it requires a significant investment of resources from individuals, from research institutions, and from industry or government. Although the Web Science and the Internet Science communities were formed mostly in a top-down fashion, the sustainability of those communities was ensured by research funding from key research institutions, national research councils, the European Union, and significant effort by individuals.

Use of a lingua franca

Beyond community formation, there are differences in the language of discourse (the lingua franca) that is employed in each area. Network scientists initially shared graph theory as their lingua franca but have more recently employed models taken from physical processes (percolation, diffusion) and game theory (Easley & Kleinberg, 2010) to describe processes on graphs. They have also moved from descriptive network metrics to the development of novel inferential techniques to test hypotheses about the evolution of a network based on various self-organizing mechanisms (Robins, Snijders, Wang, Handcock and Pattison, 2007; Snijders, Van de Bunt and Steglich 2010). As a result, the use of graph theory is not necessarily the foundation for contemporary Network Science research. Further, there is use of complex systems analysis to deal with phase changes and discontinuities between different operating regimes; these are used to study why epidemic and pandemics spread globally. As a result, many Network Science publications are featured in journals such as *Nature*.

The Web Science community has not yet embraced a lingua franca per se but one can argue that an understanding of Web standards, technologies and models (HTTP, XML, Javascript, REST, models of communication, ontologies) and of frameworks of social theory are components of what could develop into a lingua franca. The W3C has been fostering a significant part of the discussion on Web protocols and their implications. A basic understanding of the evolution of the Web on both the micro and the macro levels is the foundation for Web Science research.

Similar means of discourse are employed in the Internet Science community. For Internet Science, the components of the lingua franca include the set of Internet standards (RFCs) and associated commentary and implementation (or even C code) as in Stevens' books (Stevens, 1993; Stevens & Wright, 1995), as well as the existence of de facto standard implementations of systems in open source. They

[1] http://www.internet-science.eu

also include a basic understanding of the principles of Internet protocols, infrastructure (routers, links AS topology), social science (preferential attachment models), law and policy.

Research methodologies
In Network Science, research methodologies involve network modelling and network analysis (Börner et al., 2007; Carrington, Scott, & Wasserman, 2005) on networks that include, but are by no means restricted to, the Web and the Internet. In Internet Science, methodologies that employ measurements of engagement of Internet users with online resources and the Internet of Things are prevalent. In Web Science, mixed research methods that combine interpretative and positivist approaches are employed widely to understand the evolution of the Web based on online social network datasets, clickstream behaviour and the use of the Web of data.

Beyond methodologies, the Web Science community is working on providing the Web Science Observatory (Tiropanis *et al,* 2013, Tiropanis *et al,* 2014), a global distributed resource with datasets and analytic tools related to Web Science. Similarly, the EINS project is working on providing an evidence base for Internet Science research. And the Network Science community have a long tradition of making canonical network data sets available for use by the community along with network analysis software such as UCINET (Borgatti, Everett and Freeman, 2006) and large scale repositories of network data such as SNAP (Leskovec, 2011).

Clearly there is an overlap in the research methodologies of these three areas:

- they draw on data gathered from social networks, infrastructures, sensors and the Internet of Things,
- they involve measurement, modelling, simulation, visualisation, hypothesis testing, interpretation and exploratory research, and
- they use analytical techniques to quantify properties of a network (abstract, virtual or real) as well as more qualitative techniques.

So far, there has been significant emphasis on the social sciences in Web Science, on both social science theories and methodologies in Network Science and on protocols and computer science in Internet Science. However, these foci will change in the future and the research methods or data will continue to mingle between these three areas. For example, data on the Internet of Things might not remain exclusive to Internet Science since those data could be combined with data on human behaviour on the Web from the Web Science perspective or to explore emergence and outcomes of the networks that they enable from the Network Science point of view. Similarly, data on the behaviour of users on the Web will be used to explore the use of bandwidth in the underlying Internet Infrastructure. The different types of measurement point to the fact that often, part of the research, especially in the top-down-formed areas of Internet Science and Web Science, are associated to specific goals.

Given this shared pool of methods and data resources, each area employs mixed methods to leverage this pool in different ways according to their research agendas as illustrated in Figure 2. Those agendas are informed by different research goals.

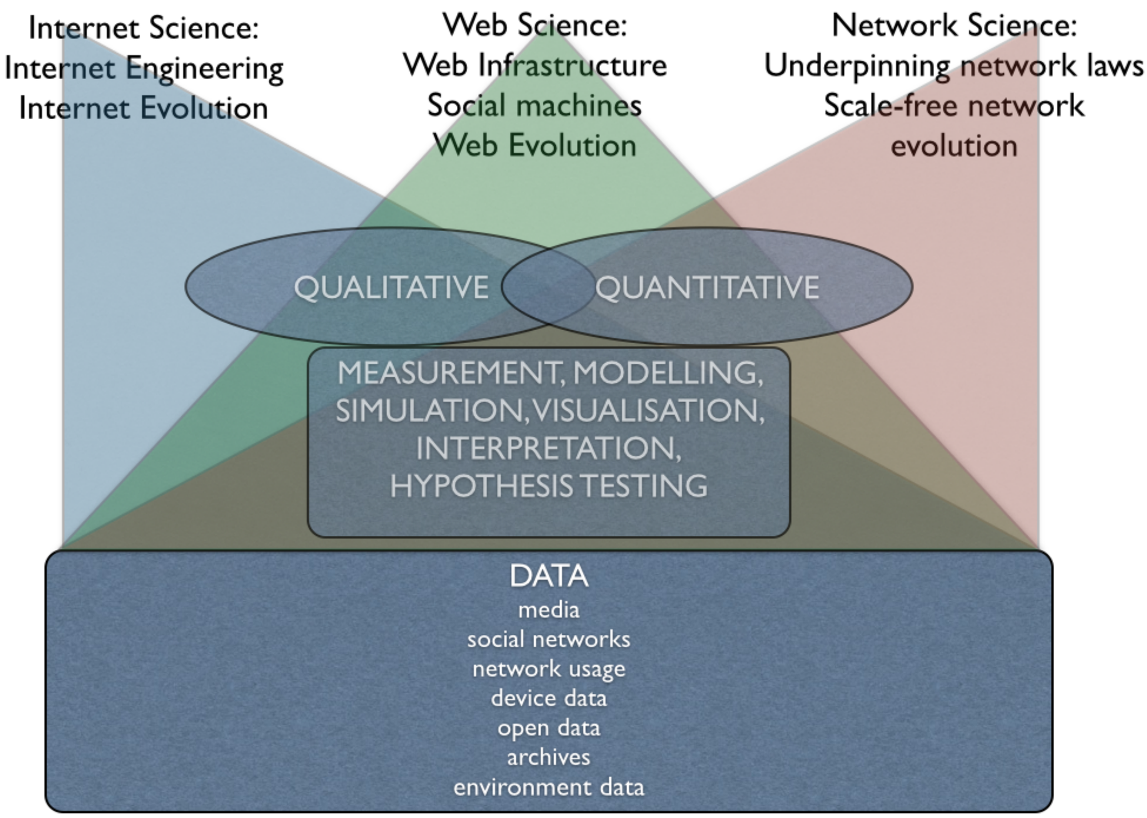

Figure 2: Network, Internet and Web Science Methodologies.

Research goals

"Web Science is focused on how we could do things better, while Network Science is more focused on how things work" (Wright, 2011); the "doing things better" refers to leveraging the potential of the Web and ensuring its continuing sustainability. Similar claims are made on behalf of Internet Science and Network Science. Although the use of the term 'science' relates to the systematic organisation of knowledge and is not directly linked to goals, we argue that goals do play a role in the formation of these interdisciplinary areas and in shaping their research agendas, scientific contribution and impact. In Web Science, the study of the Web in itself is crucial (Hendler et al., 2008), as is safeguarding the Web and its evolution (Hall & Tiropanis, 2012). In Internet Science the evolution and sustainability of the Internet and its services are central objectives; it is understood that tussles will always be the case on the Internet and that accommodating them is necessary in order to ensure its evolution (Clark et al., 2002). It seems that in both Internet Science and Web Science applied research comes first but it should be informed by the development of a basic research program. In addition, neither Web Science nor Internet Science is technology neutral; each one relies on specific protocols and standards. Further, one can argue that even the code that implements those standards embeds policy on which each respective community has reached some consensus. On the other hand, Network Science is technology agnostic and it overlaps only in part with Internet Science and Web Science, since it explores emergent structural patterns and flows on network structures be they social, biological, the Web or the Internet. Finally, Web Science and Internet Science are both also engineering disciplines; they are about building better, stronger, more robust, efficient and resilient systems. Network Science has been predominantly focused on understanding and describing emergent processes although, access to large data sets have increased interest in both predictive analytics to anticipate network changes and prescriptive analytics to optimize networks to accomplish certain desired goals. In essence, Network Science is aspiring to take insights from basic research to engineer better networks (Contractor & DeChurch, 2014).

Comparisons

Despite the differences between these areas in terms of the community formation models, lingua francas and goals, many of the research methods that they employ are common. This points to potential synergies on topics in which these areas overlap and the potential for mobilisation within those communities on topics in which there is little or no overlap. Figure 3 shows such topics from each of these areas:

1. Web Science: the area of Web-based social media is one example of primarily Web Science research. Network aspects are not the exclusive part of this since social media research focuses on associations and interaction among people and social media resources.
2. Internet Science: Research on how the Internet of Things affects information collection and transformation is primarily Internet Science research that cannot rely exclusively on network research either.
3. Network Science: Transport networks provide an example of network science research that does not necessarily relate to Internet or Web science.
4. Web Science and Internet Science: Network neutrality is an example that requires understanding of both Web and Internet technology and it does not necessarily draw primarily on network science techniques.
5. Internet Science and Network Science: Content Delivery Networks can require network techniques for distribution prediction and optimization and, at the same time, understanding of how Internet protocols and people relate to shaping that demand.
6. Network Science and Web Science: Diffusion on social media such as Twitter(™) is an example that relies on Web Science socio-technical research methods and, at the same time, on network analytic methods.
7. Web Science, Internet Science and Network Science: research on trust online or on SOPA (Stop Online Piracy Act) and its side-effects draw on all networks and on techniques that are aware of Web, and Internet protocols and code.

As the Web and the Internet continue to evolve it could be that some of the above topics will shift.

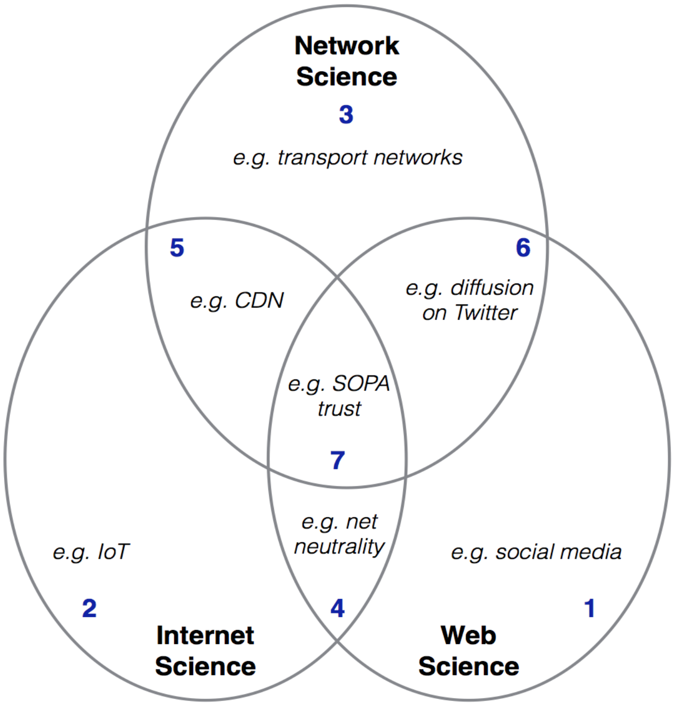

Figure 3: Research topics differentiating areas and overlaps.

CONCLUSIONS

We provide a comparison among Network, Web and Internet Science. We also propose a framework for comparing interdisciplinary areas based on their community formation, lingua francas, research methods and resources, and research goals. We can gain additional insights of the relationship among these areas by conducting co-author and co-citation analysis of publications within these areas and explore the extent to which these are distinct or merging interdisciplinary intellectual communities. Such an analysis would be even more meaningful as the related conferences and journals mature and as the similarities and differences among these areas potentially crystallise.

Both Internet Science and Web Science are technology-aware and their respective lingua francas include knowledge of the protocols and systems supporting the Internet and the Web, while Network Science is technology-agnostic. There are arguments in keeping the two layers of the Internet and the Web separate to foster innovation (Berners-Lee, 2010); consequently, Internet Science and Web Science remain two distinct interdisciplinary areas given that they have different goals, those of safeguarding the Internet and the Web, respectively. Network Science explores phenomena that include, but are not limited to, the Web or the Internet.

However, given the shared pool of mixed methods and datasets among these three Interdisciplinary areas, there are compelling benefits for collaboration to harmonise and share resources; this should be a high priority for researchers and funding agencies.

REFERENCES

Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*(5439), 509–512.

Bavelas, A. (1950). Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*. 22 (6). pp. 725–730.

Berners-Lee, T. (2010). Long Live the Web. *Scientific American Magazine*.

Berners-Lee, T. (1999). *Weaving the Web*. Texere Publishing.

Berners-Lee, T., Hall, W., Hendler, J., Shadbolt, N. & Weitzner, D. (2006a). COMPUTER SCIENCE: Enhanced: Creating a Science of the Web. *Science*. 313 (5788). p. 769.

Berners-Lee, T., Hall, W., Hendler, J.A., O'Hara, K., Shadbolt, N. & (null) (2006b). A framework for web science. *Foundations and Trends in Web Science*. 1 (1). pp. 1–130.

Blackman, C., Brown, I., Cave, J., Forge, S., Guevara, K., Srivastava, L. & Popper, M.T.W.R. (2010). *Towards a Future Internet*. European Commission DG INFSO Project SMART 2008/0049.

Borgatti, S., Everett, M. G., & Freeman, L. C. (2006). Computer software: UCINET 6. Analytic Technologies.

Börner, K., Sanyal, S. & Vespignani, A. (2007). Network science B. Cronin (ed.). *Annual Review of Information Science & Technology*. 41. pp. 537–607.

Carrington, P. J., Scott, J., & Wasserman, S. (Eds.). (2005). Models and methods in social network analysis (Vol. 28). Cambridge university press.Clark, D.D., Wroclawski, J., Sollins, K.R. & Braden, R. (2002). *Tussle in cyberspace: defining tomorrow's internet*. In: August 2002, ACM Request Permissions.

Contractor, N. S., & DeChurch, L. A. (in press). Integrating social networks and human social motives to achieve social influence at scale. Proceeding of the National Academy of Sciences.

Easley, D. and Kleinberg, J. Networks Crowds and Markets. Cambridge University Press, 2010.

Economides, N. (2008). Net neutrality, non-discrimination and digital distribution of content through the internet. *ISJLP*. 4. p. 209.

Euler, L. (1741). Konigsberg Bridge problem *Commentarii academiae scientiarum Petropolitanae* 8, pp. 128-140.

Faloutsos, F., Faloutsos, P. & Faloutsos, C. (1999) On power-law relationships of the Internet topology. ACM SIGCOMM CCR . Rev. 29, 4

Freeman, L.C. (2004). *The Development Of Social Network Analysis*. Booksurge Llc.

Frischmann, B.M. (2012). *Infrastructure*. OUP USA.

Hall, W. & Tiropanis, T. (2012). Web evolution and Web Science. *Computer Networks*. 56 (18). pp. 3859–3865.

Hendler, J. & Berners-Lee, T. (2009). From the semantic web to social machines: A research challenge for AI on the World Wide Web. *Artificial Intelligence*. pp. 1–10.

Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T. & Weitzner, D. (2008). Web Science: an interdisciplinary approach to understanding the web. *Communications of the ACM*. 51 (7).

Leskovec, J. (2011). Stanford large network dataset collection. URL http://snap.stanford.edu/data/index.html

Mitchell, J.C. (1956). *The Kalela Dance; Aspects of Social Relationships among Urban Africans*. Manchester University Press.

Monge, P.R., Contractor, N.S. (2003). *Theories of communication networks*. Oxford University Press.

Moreno, J.L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama (2nd ed.).* Oxford, England: Beacon House.

Newman, M. E. J. (2010). *Networks: An introduction*. Oxford: Oxford University Press.

Robins, G., Snijders, T., Wang, P., Handcock, M., & Pattison, P. (2007). Recent developments in exponential random graph (p*) models for social networks. Social Networks, 29(2), 192–215. doi:10.1016/j.socnet.2006.08.003

Snijders, T. A. B., Van de Bunt, G. G., & Steglich, C. E. G. (2010). Introduction to stochastic actor-based models for network dynamics. Social Networks, 32(1), 44–60. doi:10.1016/j.socnet.2009.02.004

Steen, M.V. (2011). Computer Science, Informatics, and the Networked World. *Internet Computing, IEEE*. 15 (3). pp. 4–6.

Stevens, W.R. (1993). TCP/IP illustrated (vol. 1): the protocols. *TCP/IP illustrated (vol. 1): the protocols*.

Stevens, W.R. & Wright, G.R. (1995). TCP/IP illustrated (vol. 2): the implementation. *TCP/IP illustrated (vol. 2): the implementation*.

The EINS Consortium (2012). *EINS Factsheet*. [Online]. The EINS Network of Excellence, EU-FP7. Available from: http://www.internet-science.eu/publication/231.

Tiropanis, T., Hall, W., Shadbolt, N., de Roure, D., Contractor, N. & Hendler, J. (2013). The Web Science Observatory. *Intelligent Systems, IEEE*. 28 (2). pp. 100–104.

Tiropanis, T., Hall, W., Hendler, J. de Larrinaga, Christian. (2014). The Web Observatory: A Middle Layer for Broad Data. *Dx.Doi.org*, *2*(3), 129–133. doi:10.1089/big.2014.0035

Watts, D. (2004). The ‘New’ Science of Networks. *Annual review of sociology*.

Wright, A. (2011). Web science meets network science. *Communications of the ACM*. 54 (5).