Innovation Diffusion in Scale-Free Networks : Signal Analysis in Complex Networks

Author:
Debajyoti NAG

Supervisor:
Baran CURUKLU

A thesis submitted in fulfilment of the requirements
for the degree of Master of Science

in the

Research Group Name
School of Innovation, Design and Engineering

May 2014

Declaration of Authorship

I, Debajyoti NAG, declare that this thesis titled, 'Innovation Diffusion in Scale-Free Networks : Signal Analysis in Complex Networks' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.

- Where I have consulted the published work of others, this is always clearly attributed.

- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.

- I have acknowledged all main sources of help.

- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

_____

Date:

_____

"Knowledge is the food of the soul."

Plato.

MALARDALEN UNIVERSITY

Abstract

Faculty Name

School of Innovation, Design and Engineering

Master of Science

Innovation Diffusion in Scale-Free Networks : Signal Analysis in Complex Networks

by Debajyoti NAG

There is a dire need of understanding the inner workings in a network to fully utilize the potential of the connections. This thesis aims at finding optimal parameters/conditions for spread of an idea in the society to strike a balance in utility and cost incurred for the same.

Acknowledgements

Thanks to Baran for the valuable guidance and patience.

Contents

List of Figures

List of Tables

Abbreviations

BA	**B**arabási **A**lbert model
CG	**C**ontinuous **G**rowth
PA	**P**referential **A**ttachment
CPS	**C**yber **P**hysical **S**ystems
M2M	**M**achine to **M**achine
IoT	**I**nternet of **T**hingss

For/Dedicated to/To my. . .

Chapter 1

Introduction

1.1 Networks

1.1.1 Computer Networks

Computer Networks are networks of interconnected devices that can exchange and share information and resources. These connections may be categorized depending on the form of connections being formed, and rules governing communications within the network.

1.1.2 Social Networks

Social Networks refer to arrangement to individual agents, most commonly humans, such that these arrangement are dependent on the actions of the agents and also influence their future actions. This is also referred to as a 'social structure'.

With personalization of technology, computer networks are shaping themselves to get further closer to the social networks in terms of structure, shape, size, degree distribution etc. A comprehensive review of the same was done by Newman [1] .

This is particularly important and interesting with the emergence of personalized wearable devices, which could lead to a 1:1 man to machine ratio, which would mean, when allowed, the personal devices would network exactly as their users.

1.1.3 Scale-Free Networks

A scale-free network is a network with the property that the fraction $P(k)$ of nodes with 'k' connections is determined as:

$$P(k) \sim k^{\gamma} \tag{1.1}$$

In general, the degree distribution of the network follows a power law.

1.2 State of the Art

Everything networks. Studies have shown networks to evolves over time to optimize functionalities, and increase longevity. The most coveted type of network, perhaps, is the kind formed naturally among varied entities, i.e., free scaling networks [2, 3].

The first model for free scaling networks was proposed by Barabási–Albert [2], hereinafter referred to as B-A model, and the works of Erdős & Rényi [4], Watts and Strogatz [5], Watts, Strogatz and Newman [6], Saramäki and Kaski [7], Yang and Jure [8], and Rycroft [9] also helped greatly in laying the groundwork.

It is prominent that existing social networks tend to scale freely and it is safe to assume that personalized devices, if and when allowed to network, will follow a similar trace.

1.2.1 State of the art in networked devices

IoT [10] - As the almighty Internet steps out of traditional computers to directly link everyday physical objects, overcoming the spatio-temporal boundations, it upgrades a part of our lives to the so called cloud. This makes our lives easier , but also raises confusion as the complexity of processes increase manifold.

CPS and M2M [11] - IoT is but a small part of a larger picture, CPS (Cyber-Physical Systems), where computational devices of all size and shape interact with each other and with everyday objects to perform complex tasks, ranging from temperature control in a modern house, to automatic detection of spread of a potentially fatal epidemic. M2M makes it possible for sensors spread over a large area to share the load of detecting varied signals, while some entirely different processing entity looks into that raw data and extracts meaningful information from it, and then a strong/special link actuates some physical entity to act based on the information just gathered.However, large scale CPS faces a major challenge due to Heterogeneous nature of network elements

1.2.2 Examples of State of the Art networked devices

1. GremlinMusic [12]

Gremlin showed a concept of interconnecting embedded devices in a whole new light. It not only allowed users to carry their music along (like every music player), but it allowed frends to connect their Gremlins and legally share music with each other. An optimization of storage, bandwidth (in p2p form), and monetary resources for the users. Analogy : take an iPod and put facebook and a free Spotify premium on it.

2. p-Cell technology [13]

   The new technology by Artemis seems promising, and could have crucial impact on the state of networked devices over time.

3. Swarm robots

   Swarm robots are a good example of how even heterogeneous entities can work together. Like the coordination between the Eye-Feet-Hand bots [14] to achieve their goal.

1.3 Motivation

Since networks are everywhere, it's important we understand their nature and working so as to exploit and utilize them to our benefit. Going back as far as the 18th century, the "Seven Bridges of Königsberg" [15] might be the most famous networking problem. Ranging from the Travelling-Salesman [16] to Graph-Coloring algorithms [17], insight into the working of networks have helped greatly in optimizing several issues.

1.3.1 Why Scale-Free Networks

Scale-Free networks are the most prevalent in nature. Hence, it is paramount to model networking of personal devices used by humans on the same topology. This model aims at gaining insight into the workings of such networks, so as to optimize the network elements and make use of their full potential. For eg., by finding the optimal positioning for self-assembling network of satellites, military installations can provide improved security at a lesser cost. Or multiple groups of Swarm networks, like groups of eye-hand-feet bots [14], can work together as a collective with an increased proficiency. Most importantly, due to autonomous nature of most network elements, these networks can be formed anywhere, water, air, or land, and can accomplish a multitude of tasks.

1.3.2 Human to Human

Human beings are innately complex yet strive for simplicity [18]. The challenge today as humans is to find, understand and explain the complex in its most simplistic form. This is important not only for businesses, but for overall development of human society to a point where communications may have the desired effect to a greater extent.

1.3.3 Future Sight

Imagine if smart phones were a bit smarter, and it synchronized the user's calender to include events from his/her best friend's public calender automatically, or if it alerted the user that the person they are supposed to meet has now entered the building.

Or imagine a world where every child has a personal monitoring device, in form of a wearable device, and as the children go to the playground in the evening, these devices are with them, now, not only they serve the purpose of monitoring the child, these devices communicate with each other and form a bond, in a most likely manner that the children form bonds, so while your son is busy playing, his device could find out about the studies he missed at school.

Networked devices could make life easier and yet more manageable.

1.4 Aim

This thesis aims to analyse diffusion of innovation in the society. The author simulates a real-world scenario where companies try to influence it's customers by spreading an idea or publicizing their products, and as an effect the customer agents, hereinafter referred to as 'agents' only, tend to lean towards the company if they are successfully , and enough, influenced. However, this campaigning propaganda incurs some costs to the company. The author tries to find a balance between the gain and cost for the company, in order to formulate a strategy which can be effectively used by the company to optimally attract customers.

Different companies have different policies towards customer involvement and making an effort towards further attracting the customer. This thesis aims to analyze and evaluate two of such policies, and to determine which policy is optimal for given circumstances.

Chapter 2

Approach and Tools

2.1 Approach

2.1.1 Network Model

It was an obvious choice to select scale-free topologies for modeling the network structure. But B-A model also proposes variants of the scale free structure depending on it's characteristics. These characteristics are :

1. Continuous Growth (CG)

   This states that the network grows continuously. i.e., at all times, new nodes are being attached to the the network.

2. Preferential Atachment (PA)

   This states the rule any node should follow while making connections. It states that the most connected nodes are most likely candidates to form a connection with. The connections are formed according to the probaility which is calculated as :

$$p_i = \frac{k_i}{\sum_{j=1}^{N} k_j} \tag{2.1}$$

   where,

   (a) p : probability

   (b) k : number of connections

(c) i : agent id

(d) N : total number of agents

Now, a scale free network may exhibit either or both of above characteristics. However, the author decided to include both attributes in his model to make it as close to real life as possible.

2.1.2 Orientation and View of the simulation environment

Initially, an object oriented view was used to give better control over the network agents. But this led to higher time complexity, and the author decide to switch to a connection-view model, where more focus was given to the connections being formed and everything was managed from that view. This resulted in significant decrease in time complexity. Also, focusing on the connections was easier as the whole network could be minimally represented by using the edge list.

2.1.3 The Model

The author decided to focus on a competitive diffusion where the companies compete over a single attribute of the agent. The attribute *color* was chosen for the simulation as it is easy to understand and visualise, and let's us present the effect in a 2-dimensional model where clustering is not directly dependent on the axial values. As a limiting case, the author decided to simulate only two companies. This limits the scope as it does not give much insight into cases where two companies might collaborate for mutual benefit, like to win against a third rival company.

2.1.4 Presentation Approach

Like for any project with a large number of agents, a clear and easy to understand presentation of the simulation was imperative. Initially the author went with 2-D histograms depicting the membership values of each color group, which presented nicely how the members moved around as the height of the bars representing the number of members in a group could be easily seen to rise or fall, and the speed of this descent or ascent would tell about the rat of success or failure for a particular color. An example could be seen in figure 2.1.

However, with the progress of the model this model became obsolete as it was unable to represent the aspect of "Cost" to a company. To address this necessity, the author
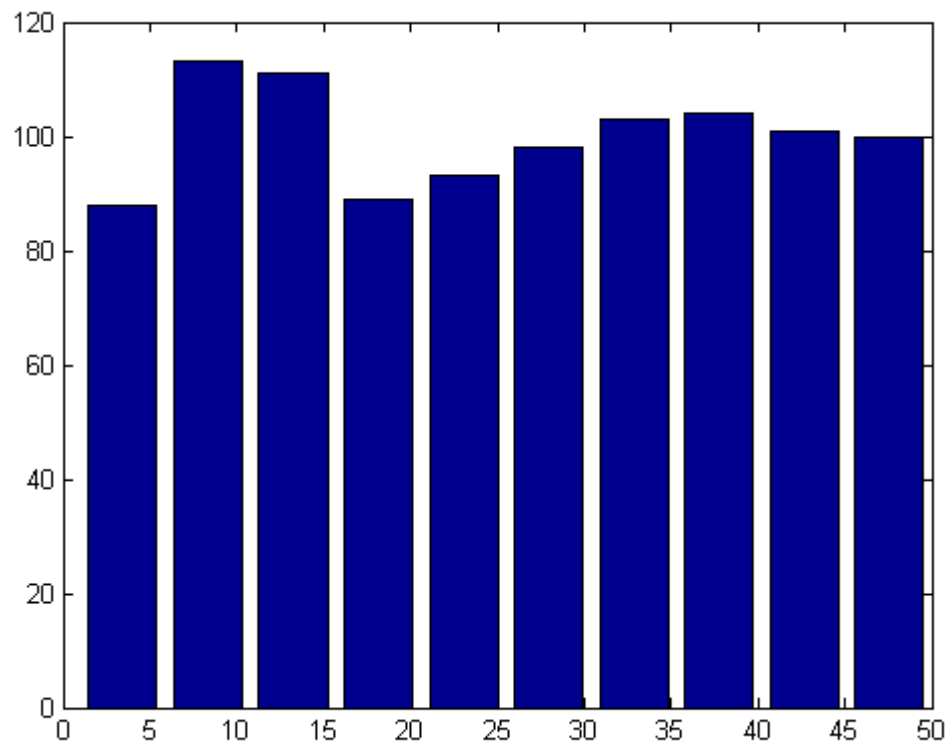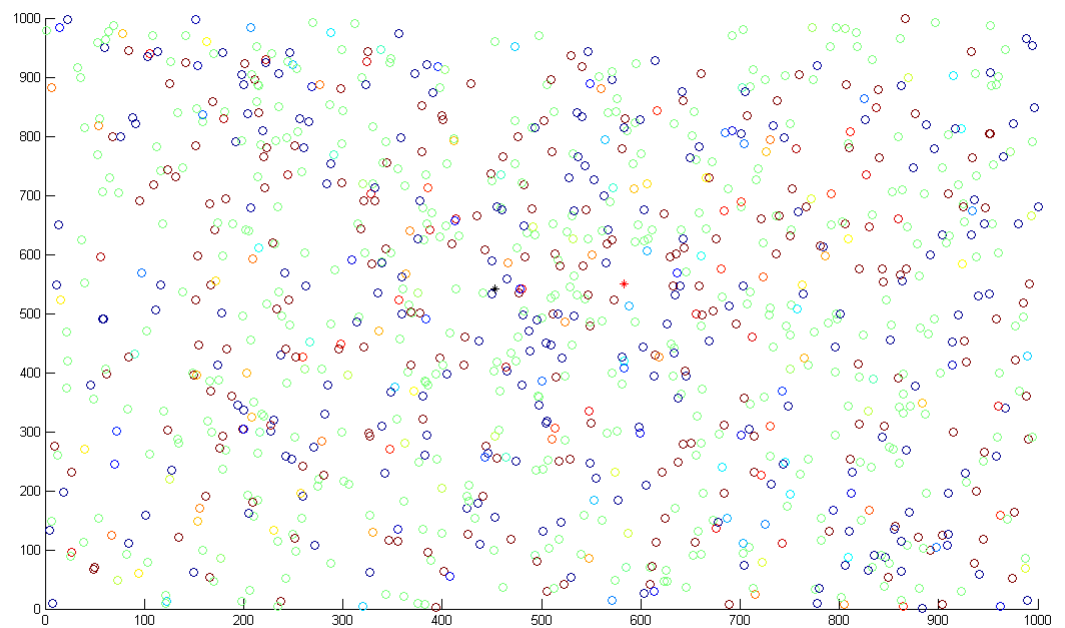
FIGURE 2.1: Color Clusters

FIGURE 2.2: 3-D Spatial Cluster

decided to switch to a 3-Dimensional presentation, with X- and Y- axes representing generic movement for agents, and the movement or effort as well as incurred cost for the companies, ad the color, being the 3-rd dimension of the presentation, depicting the value of the attribute. The speed of movement here serves the purpose as the speed of rise/fall in the previous model. An example of this is figure 2.2.

While the X-Y movement in this model, for the agents, is a correlation, the changing colors as they move represent the agents leaning towards or away from a particular company.

2.1.5 Company Policies

Different companies may have different policies towards customer involvement.

The customer involvement continuum followed by the author is influenced by the works of Bodil Sanden [19] and follow the categorization by Ives & Olson [20]. These categories are :

1 No involvement. Users are unwilling or not invited to participate.

2 Symbolic involvement. User input is requested but ignored.

3 Involvement by advice. User advice is solicited through interviews or question-naires.

4 Involvement by weak control. Users have sign-off responsibility at each stage of the system development process.

5 Involvement by doing. A user as design team member or as the official liaison with the information system's development group.

6 Involvement by strong control. Users may pay directly for new development output from their own budget or the users' overall organizational performance evaluation is dependent on the outcome of the development effort.

Out of these, this thesis focuses on categories 2 and 5. These two specific categories were chosen because they represent most dissimilar choices while still avoiding the extreme conditions, which allowed the model to simulate the more probable scenarios.

2.1.6 Simulation Approach

The simulation in this thesis focuses on companies attracting agents towards their cause, and the cost incurred in terms of efforts on the part of the company. The model witnesses three types of potential interactions :

1 Company to Agent interaction

These interactions are predefined, in a sense that a certain number of agents already seeded to believe in a company at the beginning of the simulation. This is explained further in Chapter 3.

2 Agent to Agent interactions

These interactions are the means by which one agent may influences another to lean towards a company. In this model, the agents belonging to either of the companies in question are referred to as *influencing agents*, and the others are referred to as *influenced agents*. This also mimics real life as the connections betweens agents are directional.

3 Agent to Company interactions

Whenever an agent is influenced, during the simulation, this also causes the company to make some effort, the amount of which depends on its policies. These interactions are what cause the cost to the company.

The outcome of these interactions are measured by the following two values

a Gain

This refers to a success for a company in influencing an agent to lean towards itself.

b Cost

This refers to the cost incurred to the company and is a fucntion of the effort taken by the company. It depends on the policy being followed by the company.

2.2 Tools

As a requirement from the University, Matlab was chosen as the programming language, and no special toolboxes were used. The curve fitting app was used occasionally to check the output of the simulation.

The formulas, and variants thereof, used by the author were sometimes first tested as a prototype in Python with NetworkX and Mathematica for mathematical validation. However, any of those implementations do not directly contribute to the result.

Chapter 3

Methodology

On an abstract level, the model consists of an environment comprising of numerous agents, and a few companies. This thesis specifically focuses on the various strategies used by the companies to attract the agents, and try to measure its efficiency in terms of the related costs and benefits. Based on these measurements, a company could determine the optimal strategy to maximize its efficiency for the targeted section or sub-subsection of agents.

3.1   The Companies

The companies only exist in a superficial form in the model. In this model, each company plays the role by providing a central point for the agents to cluster about, and the movement of this point itself denotes effort made by the company, which in turn incurs a cost to the company.

The author decided to keep this model limited to two companies competing for the same spot, i.e., any agent cannot totally belong to both companies at the same time. The primary reasons for this were limitations of time and computation power, in that order.

3.2   The Agents

The agents in this model are simpler constructs, each having a "color" attribute, and this attribute changes as the agents tend to believe in either of the companies.

Structurally, each agent has following attributes :

1. Id

2. x-position coordinate

3. y-position coordinate

4. Color value

5. Influence value

During its lifetime, an agents has constant value for it's Id, but the rest of it's attributes may change. The influence value of the agent is determined as a function of the number of connections it makes, and remains constant once the whole network is created. All other attributes, however, may change throughout the simulation.

3.3 Connections

The connections between the agents form in such a manner that the model satisfies the requirements of free-scaling. Following in the footsteps of the giants, the implementation is based on algorithm 1, which is a modified form of the B-A algortithm.

3.3.1 The World

The model is first built based on two fixed parameters, and then the simulation runs on it. The author starts building the model for a fixed number of total agents N, and predefined average connectivity d in the environment. The initialization of the model begins with the creation of a small, fully connected network, and then, with time, new agents are added to this network, and these agents form the connections following the rules of PA. The exact method for forming this connection was modified by the author, so as to further trick the environment and simulate scenarios of chances of connection being formed, but only a fraction of those connections actually being formed. This was achieved by employing a 'weight' modifier on the probability of an agent to form a connection. This weight, in turn, is defined for every possible pair of agents, and is based on the difference in the value of a special attribute "likemindedness", defined for each agent in an uniform, pseudo-random manner. This attribute is not considered an essential attribute because it may or may not be used, in various simulations. However, as an effect of using this modifier, the number of connections for an agent comes down considerably, while the nature of the connections still remain same, and the network remains scale free.

3.3.2 The Simulation

The simulation begins with allocating "influence" value to each agent, as a function of the number of connection it makes. This makes sure that the most-connected agents get a chance to be more influential. After this, values are uniformly assigned to all agents for the "color" attribute, and then they are seeded to belong to different companies. 20 % was the value chosen by the author for initial amount of seeding, for the following combination of reasons:

1. It allows for equal amount of seeding in both companies.

2. It allows for double amount of seeding in a neutral company.

3. After above mentioned allocations, it leaves equal amount of distribution randomly, to simulate chance.

The above rules give us the formula

$$company_1 + company_2 + company_{neutral} + company_{random} = 100$$
$$\Rightarrow x + x + 2x + x = 100\%$$
$$\Rightarrow x = 20\% \tag{3.1}$$

Then, the simulation is run, where in each time step, an edge is selected randomly, and an exchange could happen if the FromNode of this edge belongs to one of the two companies. Here, "belonging to company" refers to the value of the "color" attribute for the agent. After this, the influence value of the FromNode is subjected to a randomly selected threshold, and upon satisfaction, the actual communication occurs. The threshold is lowered depending on a company's efforts towards attracting the customer.

The selection of the edge depicts existence of a connection, while satisfaction of the threshold depicts the actual communication. This constraint was included to make the simulation model even closer to real-life. Hence, a higher effort would result in more cost, but would give a company more chances of a successful communication.

For every successful communication, the agents shift their "color" attribute a fraction towards the influencing company. This shift also translates to their x-y coordinate, which is shown in the presentation. The important aspect here is that these changes in color and x-y coordinates, while not always causally related, are always correlated.

For the companies, however, it has different significance. Whenever an agent has moved closer, the companies also make an effort and move some fraction towards the agents. The amount of distance moved by the company represents the effort it makes, and is determined by the company policies. These movements also incurs the cost to the company, and a combined function of the gain and this cost determines the effectiveness of the policy.

Towards the end of the campainging, a company may increase their effort to maximise chances of gain. This is achieved in the model by doubling the shift amount whenever a certain amount has been spent on the campaigning.

3.4   Definitions

1. *Fully Connected*

   It follows the standard definition that each node is connected to every other node. The initial network is formed in a manner such that it is fully connected and satisfies the average number of total connections.

2. *Likemindedness*

   It is a special attribute of the agents, that is used in this model to modify the probability of formation of a new connection between two agents. It simulates the real life phenomenon that people with similar interests may be more likely to form a connection.

   This subject is very vast in iteslf, and covers a wide ground [21, 22]. This is also helpful in showing the aspect that some people may be introverts or extroverts. However, that aspect is not simulated in detail in this model.

3. *Weight*

   The weight is the difference in likemindedness of two agents. It is calculated as

   $$weight_{j,k} = abs(likemindedness_j - likemindedness_k) \tag{3.2}$$

4. *Company Centre*

   The company is not implemented as a structural entity, rather, the company centre is used to keep track of the attitude of the agents towards the company, and also to track the effort and cost on the part of the company.

5. *Influence*

   This attribute determines the chances of an agent to influence another agent. As in real life, we allow more connected agents a chance to be more influential, but ultimately, there is no deterministic end to the model, and whether or not an actual comunication takes place and an agent gets influenced is always determined by a randomly determined threshold. This is calculated as a function, described by equation 3.3, of the number of connections the agent makes.

$$influence_i = \frac{connections_i}{connections_{max}} \tag{3.3}$$

6. *Shift*

   This shift represents a change in the state of the overall system. For an agent, a shift in it's color attribute represents the agent leaning towards a company. This shift is correlated with the same ratio of change in the agent's x-y coordinate, used for visualization purposes. For a company, the shift only occurs in it's x-y coordinate, which is caused by the x-y shift of the agent, and this change represents two things :

   [a] The speed and frequency of movement represent the effort taken by the company

   [b] The distance moved represent the cost incurred to the company

3.5   Algorithms

Algorithm 1 Create Scale-Free Network

 N : Input, total number of agents
 D : Input, average number of connections
 if $D < N + 1$ **then**
   Form a *fully connected* network with D+1 agents
 end if
 implement CG
 $i \leftarrow D + 1$
 while $i < N$ **do**
   $Agent_i \leftarrow newAgent$
   assign *likemindedness*
   implement PA
   $k \leftarrow 0$
   while $k \leq \frac{D}{2}$ **do**
     calculate $probability_{basic}$ based on equation 2.1
     compute $weight$ based on equation 3.2
     $probability_{attachment} \leftarrow probability_{basic} + weight$
     $Agent_i$ forms $connection_k$, based on $probability_{attachment}$
   end while
 end while

Algorithm 2 Simulation

---

assign $centres_{company}$
assign $effort_{company}$
calculate $influence$ based on equation 3.3
assign $color$ to all agents
seed 20% agents to $color1$
seed 20% agents to $color50$
seed 40% agents to $color25$
seed $2 most\ influential$ agents to different companies
J : number of iterations
$i \leftarrow 0$
while $i \leq J$ **do**
   select $edge_i$
   if $FromNode_{color} = color1$ OR $color50$ **then**
      select $threshold$
      $threshold = threshold - effort_{company}$
      if $FromNode_{influence} \geq threshold$ **then**
         $ToNode_{color} \leftarrow ToNode_{color} + shift$
         $ToNode_{x-coordinate} \leftarrow ToNode_{x-coordinate} + shift$
         $ToNode_{y-coordinate} \leftarrow ToNode_{y-coordinate} + shift$
         $Company_{x-coordinate} \leftarrow Company_{x-coordinate} + shift$
         $Company_{y-coordinate} \leftarrow Company_{y-coordinate} + shift$
      end if
   end if
end while

Appendix A

Appendix Title Here

Appendix goes here.

Bibliography

[1] M. E. J. Newman. The structure and function of complex networks. *SIAM RE-VIEW*, 45:167–256, 2003.

[2] A.-L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286:509–512, October 1999. doi: 10.1126/science.286.5439.509.

[3] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Reviews of Modern Physics*, 74:47–97, January 2002. doi: 10.1103/RevModPhys.74.47.

[4] P. Erdős and A Rényi. On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pages 17–61, 1960.

[5] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 1998.

[6] M. E. J. Newman, S. H. Strogatz, and D. J. Watts. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E*, 64:026118, Jul 2001. doi: 10.1103/PhysRevE.64.026118. URL http://link.aps.org/doi/10.1103/PhysRevE.64.026118.

[7] Jari Saramäki and Kimmo Kaski. Scale-free networks generated by random walkers. *Physica A: Statistical Mechanics and its Applications*, 341(0):80 – 86, 2004. ISSN 0378-4371. doi: http://dx.doi.org/10.1016/j.physa.2004.04.110. URL http://www.sciencedirect.com/science/article/pii/S0378437104005503.

[8] Jaewon Yang and Jure Leskovec. Modeling information diffusion in implicit networks, 2010.

[9] Robert W. Rycroft. Does cooperation absorb complexity? innovation networks and the speed and spread of complex technological innovation. *Technological Forecasting and Social Change*, 74(5):565 – 578, 2007. ISSN 0040-1625. doi: http://dx.doi.org/10.1016/j.techfore.2006.10.005. URL http://www.sciencedirect.com/science/article/pii/S0040162506001879.

[10] Friedemann Mattern and Christian Floerkemeier. *From the Internet of Computers to the Internet of Things*, volume 6462 of *LNCS*, pages 242–259. Springer, 2010.

[11] I. Stojmenovic. Large scale cyber-physical systems: Distributed actuation, in-network processing and machine-to-machine communications. In *Embedded Computing (MECO), 2013 2nd Mediterranean Conference on*, pages 21–24, June 2013. doi: 10.1109/MECO.2013.6601317.

[12] Peter. Rojas. Musicgremlin: feeding your mp3 player wireless downloads after midnight, 2004. URL http://www.engadget.com/2004/07/13/musicgremlin-feeding-your-mp3-player-wireless-downloads-after/.

[13] Artemis. Steve perlman's pcell, 2014. URL http://www.artemis.com/press.

[14] Swarmanoid. Swarmanoid: Towards humanoid robotic swarms, 2011. URL http://www.swarmanoid.org/.

[15] Eric W. Weisstein. "königsberg bridge problem." from mathworld–a wolfram web resource., . URL http://mathworld.wolfram.com/KoenigsbergBridgeProblem.html.

[16] Eric W. Weisstein. "traveling salesman problem." from mathworld–a wolfram web resource., . URL http://mathworld.wolfram.com/TravelingSalesmanProblem.html.

[17] Eric W. Weisstein. "vertex coloring." from mathworld–a wolfram web resource., . URL http://mathworld.wolfram.com/VertexColoring.html.

[18] Bryan Kramer. There is no more b2b or b2c: There is only human to human (h2h). socialmediatoday, 2014. URL http://socialmediatoday.com/bryan-kramer/2115561/there-no-more-b2b-or-b2c-it-s-human-human-h2h. [online] Accessed: 2014-05-10.

[19] Bodil Sandén. The customer's role in new service development. 2007. URL http://urn.kb.se/resolve?urn=urn:nbn:se:kau:diva-785.

[20] Blake Ives and Margrethe H. Olson. User involvement and mis success: A review of research. *Management Science*, 30(5):pp. 586–603. ISSN 00251909. URL http://www.jstor.org/stable/2631374.

[21] C.G. Jung. *The question of the therapeutic value of "abreaction"*. 1921. URL http://books.google.ca/books?id=BJo1cgAACAAJ.

[22] Joshua Wilt and William Revelle. Extraversion, 2008.