IT/CS Job Trend Analysis of the Graduates of Universities/Institutes with Leading Software Houses and Industries: Complex Network Analysis

Nadeem Mahmood[†], Muhammad Faizan[†], Jibran R. Khan[†], Kashif Rizwan[††], Farhan A. Siddiqui[†], Adnan Nadeem[†††]

[†]Department of Computer Science, University of Karachi, Karachi, Pakistan.
[††]Department of Computer Science, Federal Urdu University of Arts, Science & Technology Karachi, Pakistan
[†††]Faculty of Computer and Information System, Islamic University of Madinah, Madinah, Saudi Arabia

Summary
Everything is connected in one way or another in this universe and forms a network such as the internet, biological network (genome network, neural network), physical networks (infrastructure based networks of roads, trains and other communication networks), plant and ecological network, social networks (friendship, coworker, coauthor etc.). Complex network analysis provides us the mechanism where we built networks by identifying relations among various nodes to discover hidden characteristics of the problem domain which otherwise can't be discovered. In this paper, we investigate by formulating the network of job pattern of software graduates of leading universities/institutes of Karachi in various government departments, software houses and IT industry in Karachi, Pakistan. Further, we will evaluate the network behavior with help of network centrality matrices. This research will guide us through the industry trend and we will be in a better position to give or present a model for the growth of IT industry in Pakistan. The survey was conducted to get the findings based on the questions we asked participants and data collected from different sources of varying ages and experience. We plotted histograms of degree distribution and found that there is a possible relation between power-law and degree distribution of companies. Average job span turns out to be less than 2 years. Moreover, the analysis proposed that graduates of public universities are creating more jobs in the market, rather than private universities.

Key words:
Network science, complex network analysis, network centrality metrics, power law, degree distribution.

1. Introduction

Network science has emerged as a very important field to visualize and analyze the various complex problem domains in the form of network structure. The origin of network theory is graph theory presented by Euler[1]. The network structure is used to represent interactions among various components of the complex systems in terms of nodes and links. Its roots lie in computation and statistics. The analysis of complex network helps to identify hidden relationships and characteristics which may not be observed/analyzed otherwise[2]. The topological and structural relationships may further be investigated by applying network metrics. The application of network science in almost all domains of life including, finance, management, biology, medical science, physical infrastructures, sociology and other disciplines[3].

Network Science is the emerging field over the last couple of decades. It's a tool that gives mathematically support from graph theory to analyze, calculate & model network structures and graph concepts[4]. We want to study the pattern of job trend of a software developer's hiring in Pakistan/Karachi IT industry and will evaluate the network behavior with help of matrices provided by Network science. It has been observed that software professionals switch their jobs very often due to different factors such as social, economic & personal interest and due to the number of options they have in software houses[5,6]. We will try to understand their switching patterns. It is also interesting that how entrepreneurs have impacted the industry in Pakistan & what is the influence of Institutional background of getting employed in a certain reputed company on a certain degree of designation. This research will guide us thru the industry trend and we will be in a better position to give or present a model for the growth of IT industry in Pakistan.

1.1 Network Characteristics and Topology

1.1.1 Regular Graph

Regular Graphs are graphs with the same degree on each node, if it's a di-graph then each node's in-degree & out-degree should be equal, shown in the following Figure 1.

Manuscript received January 5, 2019
Manuscript revised January 20, 2019

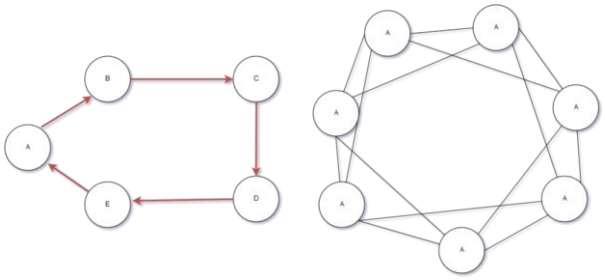

Fig. 1 Regular Graph

1.1.2 Bipartite Graph

Bi-partite graphs are such graphs that have two sets of vertices, and one condition that must be fulfilled is that a set of the vertex can only connect to another set of vertex and are not interconnected as shown in Figure 2.

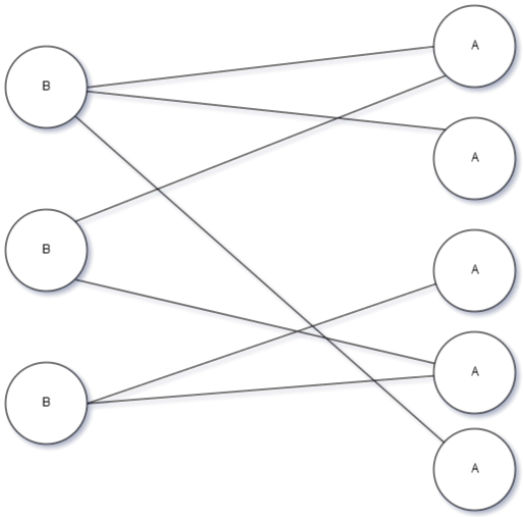

Fig. 2 Bipartite Graph

2. Research Methodology

2.1 Survey Methodology

The link was spread in social networks such as Facebook and LinkedIn. On Facebook it was sent to close network of those who are related to IT field somehow, the responder of the survey can be any professional working related to IT field, be it developer, network administrator, Data scientist in a software house, Corporate IT sector, directors and executive of software industries[7,8].

The questionnaire contains 4 sections, first sections contains demographic questions, that includes, name, email, age-range, gender, the second section contains information related to education, this section requires questions such as last educational achievement (degree), when it was completed, what was the institute and its type either it was private sector or public. The third section asked about the career, like current, previous and first job, experience on job, type of industry, years of experience, etc. the fourth section asks about online existence of the person, it inquires about the research and educational network a person is connected to, about the social networks and professional network[9].

2.2 Data Collection Tool

We have used Google Form to receive the data. Also, the response sheet was shareable, this allowed to share my data without letting it be editable. Also, it created an embedded link allowed to be posted anywhere across the web and social network. Through the survey, a collection of 100+ records were made, which were from the target audience, they were IT professional and had several years of experience, next the graph will be formed on the basis[10,11].

2.3 Analysis and Network Formation

The graph of job network is amazing it gives us an insight of how companies hire and have preferences on hiring from certain institutes. It is also critical to analyze what type of organization a person is likely to be in depending on from where he graduated. Because during analysis. There was a pattern that some companies form a close bond while other are quite isolated. We will further discuss this pattern in a later section.

2.4 Network Formation

It was needed to analyze that how an institute improves your chances to land in a certain type of organization, we needed to have certain parameters, such as a set C represents the companies and their related information, while other set I represented the institute or education center. Now there is a connection of what type of company is hiring a certain person and from which institute, although the selection in a company relies on person individual intellect, but institute have certain patterns and companies have their preferences that can be seen and studied[12-15].

Having known the set of problem, the network topology best maps on two-mode network or Bi-partite formation of the network. A two-mode graph needs three parameters. Two set of nodes and a set of edges that links one set to another. So we can represent the two-mode network as a function of $T(P, S, E)$. Where T is a function of Graph G which takes input parameters P as a set of primary nodes, S is a set of secondary nodes and E represents the links between these sets.

2.5 Data Formation in Network

The data is formed in following order and dimension so as to bring the best result out. We have following data formation. Where each node information are presented under their respective heads, see Table 1-3.

Node Companies C

NA node represents the companies that hold employees or someone is working, please be noted that this doesn't include all the companies, the nodes are based on a very small set of data collected from a local area and over the internet using Google Form.

Table 1: Company node information

Attributes	Description
Id	Id uniquely identifies a node within a graph. It's numeric in value
Label	The label is the company's name it is used to display name on the graph if needed for the purpose
Employees Head	Employees head is a range of employees an organization may possess. This data is extracted from the Linkedin profile of the company itself. This gives an idea of how big an organization is
Industry Type	Industry type is an important aspect of analysis, we wanted to figure what kind of industry attracts certain students of the institute, the kind of work a company does will give us more insight of where the students go after graduating from a certain school of set I.
Node Type	Node type is the classification factor in two modes, this belongs to a set of node belongs to set C.

Node Institutes I

Table 2: Institute node information

Attributes	Description
Id	Id uniquely identifies a node within a graph. It's numeric in value.
Label	The label is institute name it is used to display name on the graph if needed for the purpose.
Institute Type	Institute type tells whether an institute is private or public in set I. The private institutes belong to some business sector while public institutes are owned by public managed by the gothe vernment of the state.
Node Type	Node type is the classification factor in two modes, this belongs to a set of node belongs to set I.

Edges

In the network a link is defined over a relationship when a person is hired by a company, a connection between company and institute is created.

Table 3: Edge node Information

Attributes	Description
ID	Id uniquely identifies a node within a graph. It's numeric in value.
Label	The label is the company's name it is used to display name on the graph if needed for the purpose.
Source	The source is the point from which a person educates itself. It is from set I.
Target	Target is the point where a person gets employed to. It is from set C
Type	Type refers to the kind of edge, in the current scenario is constant and undirected
Education	Education refers to the name of degree a person attained, for example, it could be BS, BE, MS, MSC etc.
Discipline	Discipline is the Major of Degree a person attained, for example, Computer Science, Information Technology, Computer Engineering etc.
Experience	Experience is the year of experience a person have currently
Self Employed	These fields ask the employee whether he is self-employed or not.

3. Data Analysis

The data was spread on the social network over Facebook, Linkedin this caused a certain community to be more interact towards this survey. However the survey collects the data of many different universities, we also have data from different age group, and the IT industry in Pakistan is young and waiting to be mature, although it has proven itself to be quiet.

3.1 Network Explanation

As the attributes and network are already discussed.

Network component

As already discussed, the network is formed in two modes, when a person of a certain node is hired in a company link is created, noticing Fig. 3 there is a green node in the center while pink nodes are around it.

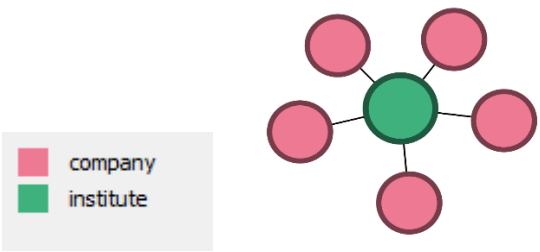

Fig. 3 Component of Two-Mode Job Network

The green nodes represent an institute, while the pink nodes are the companies in which a person is hired after graduating from that institute. The network color may alter in further illustrations as per the need of further classification of nodes.

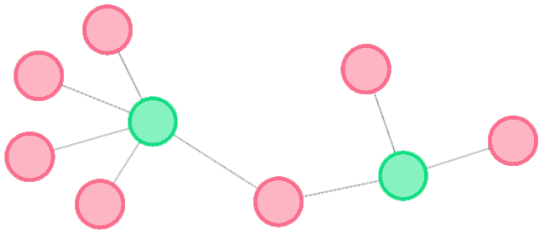

Fig. 4 Company hiring from two institute

In Fig. 4, we see an organization hiring from different institutes. This could lead us in answering questions such as, in which educational institute, a company or industry trusts? What are the chances of a person getting hired in a certain industry? Are the chances equal and carry the same weightage for some institutes? These questions can be answered well when the network projects from institute-company relation to institute-institute relation.

4. Analysis of Two-Mode

In this section, we would analyze the network in a different configuration and make statements about it.

4.1 Modularity Analysis

A complex network is difficult to study in its existing form because, with a high number of nodes, it is very difficult to analyze the structure and what is the impact of certain community and how can we find communities in the network[16,17]. Modularity is figured by analyzing strongly connected clusters and sparsely connected to remaining clusters or groups of the network; a group within a social network can be a group of friends or group of people with common interest, it can be a group of people connected to commonplace of activity such as job or education. In large networks such as web pages, groups can define web pages of the same domain or similar topic[18,19]. In this analysis, we will see the connected communities and their boundaries, as we look into the complete picture of our network. There are several things that we can know. The group is independently functioning modules we can analyze which companies are relying on which type of people, although the data set is minimal and is not a good approximation of population. But still, we would see how modularity gives us the way to know this.

The steps of finding modularity are quite complex, counting edges isn't an effective way although we know that a group will be dense within a group but sparse among other groups[20]. This is a special kind of property of network. Another way can be to categorize node so is to find the minimum number of edges going around the groups, this approach is referred to as "Minimum Cut" approach used in the literature of graph partitioning[21]. Although the graph partitioning is far away from determining community structure, and we cannot know a number of members in a community prior to finding community. Also if we put all the vertices in one group and none in the other group there would be 0 inter-community edges and thus it would be assumed minimum, sown in Figure 5 and 6.

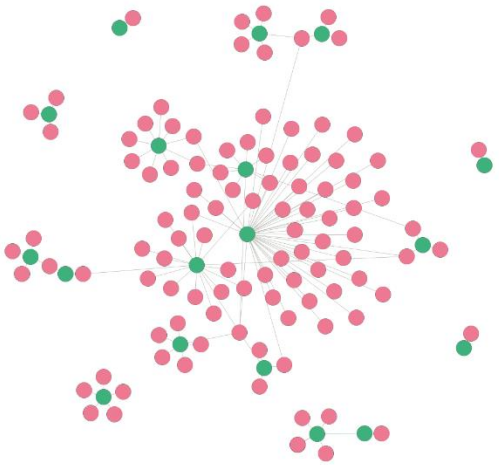

Fig. 5 Modularized View

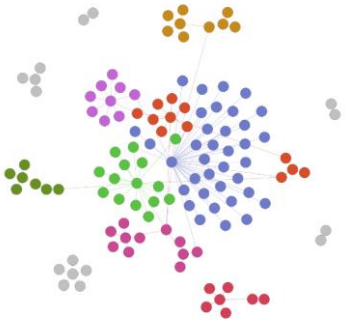

Fig. 6 Simple View

To find a community a better approach than counting edges is to observe something meaningful so that it can be evident that we have identified a community. One approach can be if the number of edges is significantly less than what we would expect to have, or if density within a group is denser

than what it could be by chance then it can be said that we have found a community.

Let's assume that a network contains n nodes, we have to divide a network in two communities, let $s_i = 1$ if a node i belongs to community A and $s_i = -1$ if it is a member of community B. Let the links between nodes i & j be given by A_{ij} (Adjacency Matrix). Meanwhile, the expected links between nodes i & j if links are randomly inserted is $\frac{k_i k_j}{2m}$, where k_i & k_i are degrees of i &j and m is total number of links calculated by $\frac{\sum_i k_i}{2}$. Hence modularity Q can be calculated using

$$Q = \frac{1}{4m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)s_i s_j \quad (1)$$

We now see the groups of companies forming a strong link with institutes. Here is the resultant comparison is shown in Table 4.

Table 4: Modularity and percent of the graph

Modularity Class	***Percent of Graph***
Blue	29.27%
Green	12.2%
Red	9.76%
Purple	8.94%
Light Brown	7.32%
Light Purple	7.32%
Maroon	5.69%
Grey-1	4.88%
Grey-2	3.25%
Grey-3	1.63%
Grey-4	1.63%
Grey-5	1.63%

We were able to identify 12 groups showed in Figure 6 that shows how an institute and group of companies are closely bonded. The dark blue component is due to the fact that the record of one institute was more than other groups. However, the type of groups is different. This shows that there are communities within industry although we have a few companies such as "Venture Dive", "Systems Limited" that has a person from 3 different Universities.

4.2 Industry-Size to Institute Analysis

The section highlights how a graduate of the certain institute is more likely to join an industry of certain size. We have data of approximate heads in a company or an organization. And we wanted how a person is likely to be accepted in an organization of certain number. Although the dataset is quite small to make this assumption. Still, we can have an idea of how likely it is to be accepted in a great organization.

University of Karachi

Graduates of the University of Karachi are great contributors to the field of Computer Software and IT. As the great institution graduates of University of Karachi lands on versatile opportunities of the job, displayed in Table 5. We wanted to figure out what is the probability for a graduate to land on a job of a certain number of people.

Table 5: UoK Employing Organization Size Probability Table

Organization Size	***Probability***
1-10 employee	0.095238
11-50 employees	0.404762
51-200 employees	0.119048
201-500 employees	0.095238
501-1000 employees	0.095238
1,001-5,000 employees	0.071429
5,001-10,000 employees	0.023810
10,000+ employees	0.095238

We can study this phenomenon in network formation, and see how different organization hire graduates of this institutes and just to give an idea we can study how great an organization is by looking at the number of employees that organization have.

Fig. 7 University of Karachi graduates hiring company's graph

As noticed in 오류! 참조 원본을 찾을 수 없습니다. we can see how different size of the company is interested in hiring graduates of University of Karachi, we can assume that the University of Karachi is a hub in the field of IT in Karachi. And the graduates are more likely to get into the organization in 201-500 employees organization.

4.3 Collective Industry Analysis

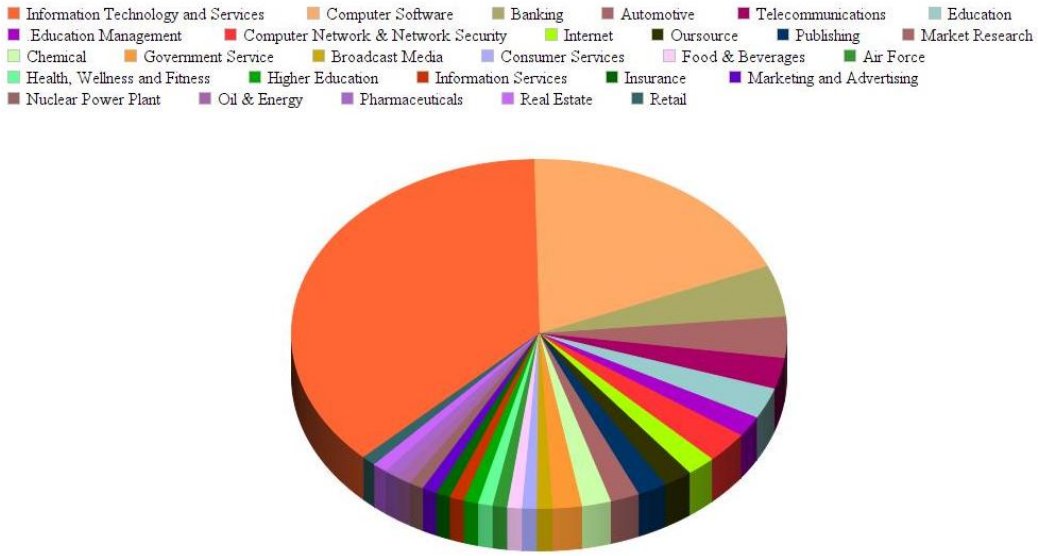

Fig. 8 Industry Acceptance Pie Chart

The fact that computer science is evolving so rapidly and spreading epidemically is because of the fact that it is highly accepted and needed in all the businesses of concern, from very small market to big business hubs the computation power helps evolve the business thus different industries approach to the computer science students in order to find the solution of this rapidly growing market and keep up with the competition[22,23]. We will analyze how different fields try to accommodate computer science graduates.
We generate a pie chart, Figure 8, which shows how many collective graduates fall into that industry type.

We can see that most of the graduates work in the IT Services that chunk is greatest in a pie chart, it is because of the industry working pattern that works on taking different projects from the international market and then providing the work after developing it in Pakistan. The second greatest chunk belongs to Computer Software Manufacturer those are the companies that work on their own software and profit by selling them internationally and locally as the product or service[24]. The banking sector is quite interestingly adopting computers era as the need of evolving economy requires fast transactions and good services. One

thing of concern here that computer networks and network security are of great importance, configuring managing networks and its related work are required by corporates and different local industries to manage their business and secure their data however within the sample data we see that such companies are lower rather industry hires an employee to manage their networks. There is various another industry type that can be observed in the above charts[25].

4.4 Job to Experience Relation

In this section, we will concern how long a person works in a company, how long they are likely to work. Because we noticed that people have started to switch job more frequently than before, that is because of various reasons in Figure 9.

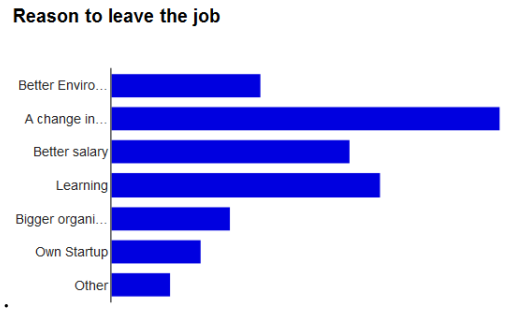

Fig. 9 Reason for leaving the last job

We gave them options to select from, 17% left their job for "Better Environment" 46% were looking for "Change in career" another two great reasons of leaving a job were for "Better Salary" & "For Learning" with percentages of 28.6% and 32.1% respectively.

Fig. 10 Histogram of Experience to Job Ratio

We can see that people tend to switch job before the 2nd year of their job. Because of the reasons mentioned earlier and the result shown in Figure 9 and 10.

Power-law on Degree Distribution

We tried to find if the degree distribution of this network follows power-law, so does companies are growing or getting more familiar in different institutes with the phenomena of power law[26]. We checked the degree distribution of all the companies hiring from the different institute and created a power law plot in Figure 11.

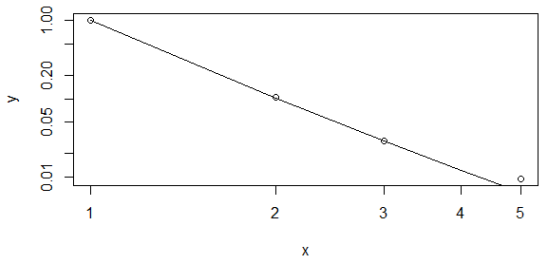

Fig. 11 Degree Distribution Power-law plot of set C

Although the graphs seem to show a good estimate the power law seems to be deviating for degree 5 that can be because of the fact that we do not have sufficient data to support or map a power law of such degree distribution, but to the initial analysis it seems that it may follow power-law distribution and can be proved with further data addition. In figure 12, the histogram 아래 we can see that company degree distribution seems to have a heavy tail. Thus there are good chances that this distribution will follow power law as the data grows. This power law has $\propto = 3.627503$.

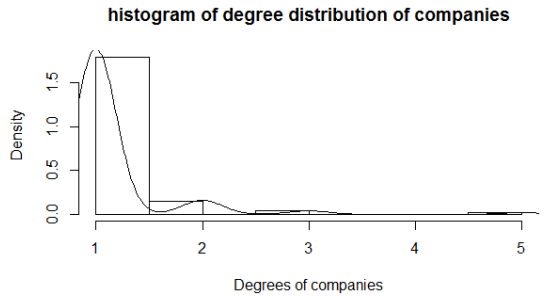

Fig. 12 Histogram of Degree Distribution of Network set C

5. Conclusion

In this work, we first looked at the graph and its basic structures, topologies, and matrices inherited by network science, and then we looked at properties and matrices that network science have evolved due to the fact that researchers were applying on a different set of problem, for example, between-ness was evolved by social network analysts. Because they found it to be important characteristics of a network's component. We also discussed the topologies of the network, the motives behind forming a network and the challenges faced in complex systems and how the network is helpful in analyzing the

entire and component-based working of the system, and why is it easier to understand and manipulate.

We looked into the methodology of research. The survey was conducted to get the findings based on the questions we asked participants and data collected from different sources of varying ages and experience. Our formation and observation of this network analysed greater picture of the industry of Information Technology, and its relationship with local education hubs, we figured the impact of educational institute over the local industry with modularity calculation and found out that certain group of companies was hiring more from a specific institute, We inspected the various kind of industry and how saturated industry is for a computer science graduate, it was found that they are highly accepted in a variety of industries, because of the needs of industry. The task of maintenance of industrial warehouses and network administration and many other tasks of development and IT related jobs in the industry gives the edge of getting accepted in any kind of industry from software services to publishing.

We then applied some matrices over the network and analyzed it using its values. We plotted histograms of degree distribution and found that there is a possible relation between power-law and degree distribution of companies. We recognized the great components of markets and how they are connected and disconnected on the basis of the institute they hire from.

We also figured the average job span of an employee within a company, and its relation to the company's environment and other opportunities. That turns out to be less than 2 years before a person leaves his job and gets a new one. Our analysis proposed that graduates of public universities are creating more jobs in the market, rather than private universities. We further look to enhance the data to Pakistan wide survey and IT industry, this can be considered as base layout or initial step of studying the market for any graduates, this can help to analyze acceptance of certain graduates in the industry and the major competitions among institutes.

References

[1] Euler, L., 1953. Leonhard Euler and the Königsberg bridges.. cientific American 189.1, pp. 67-70.

[2] Newman, 2010. Networks: an introduction. s.l.:OUP Oxford.

[3] Collins, F. S. a. V. A. M., 2001. Implications of the Human Genome Project for medical science.. Jama 285.5, pp. 540-544

[4] Barabasi, A.-L., 2016. Network Science. s.l.:Cambridge University

[5] Duncan Watts and Steven Strogatz, 1998. Collective dynamics of 'small-world' networks. Nature 393, pp. 409-10.

[6] Barabasi, Albert-Laszlo, and Zoltan N. Oltvai, 2004. Network biology: understanding the cell's functional organization. Nature reviews genetics 5.2, pp. 101-113.

[7] Girvan, N., 2002. Community Structure in Social and Biological Networks. PNAS 99, 11 June.pp. 7821-7826`

[8] Milgram, S., 1967. The small world problem.. Psychology today 2.1, pp. 60-67.

[9] Dietrich, J. P., 2008. isentangling visibility and self-promotion bias in the arxiv: astro-ph positional citation effect. Publications of the Astronomical Society of the Pacific 120.869, p. 801.

[10] [Faloutsos, Michalis, Petros Faloutsos, and Christos Faloutsos., 1999. On power-law relationships of the internet topology. ACM SIGCOMM computer communication review. Vol. 29. No. 4., pp. 251-262.

[11] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon, 2010. What is Twitter, a social network or a news media?. s.l., ACM.

[12] Johan Ugander, B. K. L. B. C. M., 2011. The Anatomy of the Facebook Social Graph. arXiv preprint arXiv:, p. 1111.4503.

[13] Gupta, S. X. Y. a. K. L., 2015. Structural properties of ego networks.. Social Computing, Behavioral-Cultural Modeling, and Prediction, pp. 55-64.

[14] Varela L.M., Rotundo G., Ausloos M., Carrete J. (2015) Complex Network Analysis in Socioeconomic Models. In: Commendatore P., Kayam S., Kubin I. (eds) Complexity and Geographical Economics. Dynamic Modeling and Econometrics in Economics and Finance, vol 19. Springer, Cham

[15] Justin Ruths1, Derek Ruths, 2014, Control Profiles of Complex Networks, Science 21 Mar 2014: Vol. 343, Issue 6177, pp. 1373-1376.

[16] Opsahl, T., 2012. Clustering in Two-mode Networks. [Online]

[17] Available at: http://toreopsahl.com/tnet/two-mode-networks/clustering/

[18] Barabási, A.-L., 1995. Dynamics of Random Networks: Connectivity and First Order Phase Transitions. [Online]

[19] Available at: Dynamics of Random Networks: Connectivity and First Order Phase Transitions

[20] Gilbert, E. N., 1959. Random graphs. The Annals of Mathematical Statistics:30, pp. 1141-1144.

[21] Adamic, L. A. a. N. G., 2005. The political blogosphere and the 2004 US election: divided they blog.. s.l., ACM, pp. 36-43.

[22] Kas, M., 2011. Structures and Statistics of Citation Networks. Ankara: s.n.

[23] John Kelly, D. F. M. S., 2005. Stanford Online Deliberation Conference, 2005. s.l., USENET Newsgroups.

[24] Moed, H. F., 2007. The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section.. ournal of the American Society for Information Science and Technology 58.13, pp. 2047-2054.

[25] Barabasi, A.L. and Jeong, H. and Neda, Z. and Ravasz, E. and Schubert, A. and Vicsek, 2012. Evolution of the Social Network of Scientific Collaborations. Physica A: Statistical Mechanics and its Applications Vol.311, pp. 590-614.

[26] Newman, M. E., 2001. Who is the best connected scientist? A study of scientific coauthorship networks.. Phys. Rev. E 64.016131.

[27] Newsgroups, U. S. E. N. E. T, 2005. Stanford Online Deliberation Conference, 2005. s.l., s.n.

[28] P. Erdos, A. Rényi, 1959. On random graphs. I. Publicationes Mathematicae (Debrecen), pp. 290-297.

[29] Strogatz, D. W. a. S., 1998. Collective Dynamics of 'small-world' networks. Nature 393, pp. 440-442.