


Nov 7th, 9:00 AM - 5:00 PM

Assessing Clustering in a Social University Network Course

Orgeta Gjermëni

University of Vlora, o.gjermeni@gmail.com

Follow this and additional works at: <https://knowledgecenter.ubt-uni.net/conference>

 Part of the [Computer Sciences Commons](#), and the [Digital Communications and Networking Commons](#)

Recommended Citation

Gjermëni, Orgeta, "Assessing Clustering in a Social University Network Course" (2015). *UBT International Conference*. 90.
<https://knowledgecenter.ubt-uni.net/conference/2015/all-events/90>

This Event is brought to you for free and open access by the Publication and Journals at UBT Knowledge Center. It has been accepted for inclusion in UBT International Conference by an authorized administrator of UBT Knowledge Center. For more information, please contact knowledge.center@ubt-uni.net.

Assessing Clustering in a Social University Network Course

Orgeta Gjermëni

Department of Mathematics, University "Ismail Qemali"
Vlore, Albania
o.gjermeni@gmail.com

Abstract. A collection of data is gathered from surveys held in a Spring Course of the Economic Faculty in the University "Ismail Qemali" of Vlora, Albania. The data set for each student contains the names of the other students through which he/she have a "social relationship". This relationship includes frequent communications, discussions on exercise solutions, and sitting usually close to each other in the class. We have constructed. At the end of the course, a final network based on this type of relationship. We are particularly interested on the clustering coefficient of this network and assessing it's "significance", in the sense of being somehow unusual or unexpected. Simulated random graph models, using R platform, are used to test the "significance" of the observed clustering coefficient.

Keywords: Social networks, Clustering coefficient, Random graph models

1. Introduction

Relationships and collective behavior in many fields are popularly studied as part of network analysis [11 – 13]. Real world networks can be seen as complex systems, which are unstable, multi – parameter, nonlinear and time varying. Information flows about jobs, new products, technologies or political opinions, spread of epidemic diseases are some examples of social networks [1, 2] effects in our lives. The concept of clustering in social networks was formally introduced by Watts and Strogatz [3], but also in sociology [2]. Clustering coefficient, known also as transitivity, as a global measure, summarizes the relative frequency with which connected triples close to form triangles. It gives answer to the question: – How well vertices in a network tend to be clustered together? Also, clustering coefficient may be seen as the probability that two nearest neighbors of a vertex are also of one another. The motivation for this interest in clustering coefficient is because in real life, our beliefs, decisions and behaviors are constantly influenced by people with whom we regularly or sporadically interact. Clustering coefficient plays an important role as a statistic parameter on analyzing those networks. For every observed network, G^{obs} , it is possible to find its clustering coefficient, $c_l(G^{obs})$, but our purpose in this study is on assessing whether it is 'significant', in the sense of being somehow unusual or unexpected. In this paper we analyze a social network, conceived as a fixed set of vertices and edges from the perspective of clustering coefficient of the network. Each vertex represents a university student, and each edge represents a 'social relationship'. The relation is defined by frequently communications, discussions on exercise solutions, or sitting usually near each other in the classroom.

The reminder of this paper is organized as follows: In Section 2, we describe the method that is applied. Next, in Section 3, we see the results, and in Section 4, we discuss about the results. Finally, we present some concluding remarks and future work in Section 5.

2. Methods

Data was obtained from surveys which were held periodically during a Spring 2015 – semester course (15 weeks) of the second year students at the Economic Faculty in the University "Ismail Qemali" of Vlora. For each semester course, students have the possibility to choose the lecturer they want

between some alternatives. A ‘mixing process’ happens at the beginning of every course within the various groups of students.

During the surveys, each of the students showed the names of other students, for which he/she had a ‘social relationship’ with. This social relationship includes frequent communications, discussions on exercise solutions, and sitting usually close to each other in the class. In this way is defined the socialization that happens within the course.

A social relationship is assumed to be ‘forever’ from the moment it starts between two students, till the end of the course. Based on the conducted surveys, a final social network graph is conceived as a fixed set of vertices and edges $G^{obs} = (V^{obs}, E^{obs})$. The orientation of the edges is neglected and the network graph is simplified.

Since for every observed network graph, it is possible to find its clustering coefficient, $cl(G^{obs})$, in our study we are interested in assessing whether this value is ‘significant’, in the sense of being somehow unusual or unexpected. Hypotheses we arise here is:

$H_0 : G^{obs}$ can be viewed as a uniform sample under either random graph model.

Network graph models are used for a variety of purposes, also for testing the significance of a pre – defined characteristic in a given network graph. Significance is defined in comparison to an appropriate form of reference. Random graph models [4] are used in setting up such comparisons. Therefore, for our purpose we will define two collections:

Γ_1 – Collection of random graphs of the same order and size as G^{obs} , classical random graphs;

Γ_2 – Collection of random graphs with the same degree distribution as G^{obs} , generalized random graphs.

Let $P(\cdot)$ be a uniform probability distribution over Γ_1 , and Γ_2 . The value of $cl(G^{obs})$ is compared to the collections of values $\{cl(G) : G \in \Gamma_1\}$ and $\{cl(G) : G \in \Gamma_2\}$. If $cl(G^{obs})$ is judged to be extreme with respect to this two collections, than that is taken as evidence that G^{obs} is unusual in having this value [5, 6]. This evidence will contradict the null hypothesis H_0 . Simulations are conducted through “vl” method [7] as a generator and Monte Carlo algorithms are used for randomizations. This is implemented in R [9], package igraph [8].

3. Results

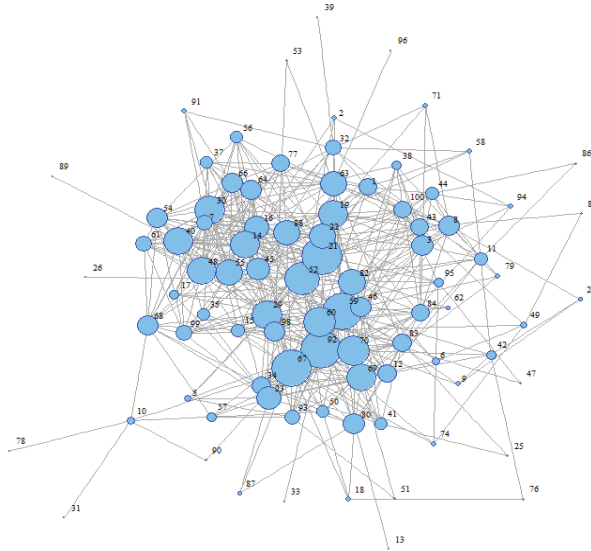


Fig. 2. This shows a visualization of the Social University Network Course, G^{obs} using layout generated by the Kamada – Kawai algorithm [10]. The size of each vertex is specified by its degree. In the simplified observed network graph were found a total of 89 vertices and 494 edges. It resulted connected and the clustering coefficient was $cl(G^{obs}) = 0.375$. A visualization of the Social University Network Course (the observed network graph) is given in Fig.1. After simulating two uniform samplings of 10.000 random graphs for each of the collections Γ_1 and Γ_2 , clustering coefficient was calculated for each graph sampled. Histograms in Fig. 2 show a summary of the resulting clustering distributions.

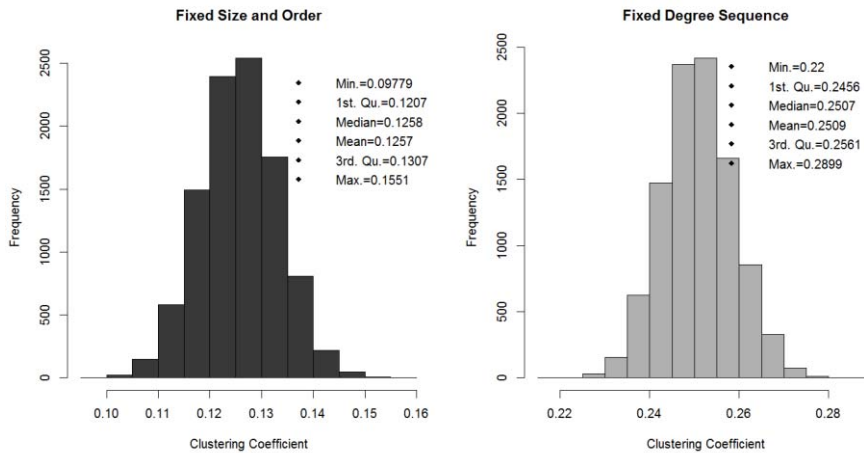


Fig. 2. Histogram distributions of clustering coefficients $cl(G)$ detected for random graphs, G generated uniformly from Γ_1 (same size and order as the observed graph) and Γ_2 (same degree distribution as the observed graph).

4. Discussion

Clustering coefficient is considered significantly high if $cl(G^{obs}) \gg cl(G)$, where G is a random graph from Γ_1 or Γ_2 . Based on the results, we can say that there is a strong evidence to reject the null hypothesis H_0 . Clustering coefficients values in both simulated collections are relatively small compared to the observed clustering network graph value.

The so called 'karate club network' of Zachary [14], has a somewhat similar nature with our network. The clustering coefficient of the 'karate club network' was 0.2556818 [15]. Compared to this value, our observed network clustering value is greater.

Conclusions

As a conclusion, at the end of this paper we can say that the observed network graph showed significant greater clustering coefficient compared to that of random graph models with a comparable magnitude (i.e., with respect to order and size) or comparable connectivity (i.e., with respect to degree distribution). We can't assume that our network graph is a uniform sample under either random graph model. Further investigations should be done on analyzing average path length and its 'significance' with the purpose detecting 'small world' [3] behavior in this network graph.

Acknowledgments go to the students of Spring 2015 – semester course of the second year – Economic Faculty at the University "Ismail Qemali" of Vlora for their collaboration on sharing personal data.

References

1. Scott, J.: Social Network Analysis: A Handbook, Sage Publications, London, 2nd ed. (2000)
2. Wasserman, S., Faust, K.: Social Network Analysis, Cambridge University Press, Cambridge (1994)
3. Watts, D. J., Strogatz, S. H.: Collective Dynamics of 'Small – World' Networks, Nature, Vol. 393. pp. 440 – 442, Macmillan Publishers Ltd (1998)
4. Erdős, P., Rényi, A.: On Random Graphs, Publicationes Mathematicae, Vol. 6, pp. 290 – 297 (1959)
5. Kolaczyk, E. D., Csárdi, G.: Statistical Analysis of Network Data with R, Use R! 65, DOI: 10.1007/978-1-4939-0983-4_5, Springer Science + Business Media, New York (2014)
6. Kolaczyk, E. D.: Statistical Analysis of Network Data with R, Springer Series in Statistics, DOI 10.1007/978-0-387-88146-1_6, Springer Science + Business Media LLC (2009)
7. Viger, F., Latapy, M.: Fast Generation of Random Connected Graphs with Prescribed Degrees, <hal-00004310>, (2005)
8. Csárdi, G., & Nepusz, T.: The igraph Software Package for Complex Network Research, InterJournal, Complex Systems 1695. <http://igraph.org/>, (2006)
9. R Core Team R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>, (2015)
10. Kamada, T., Kawai, S.: An Algorithm for Drawing General Undirected Graphs, Information Processing Letters, Vol 31(1), pp. 7–15 (1989)
11. Albert, R., Barabási, A.-L.: Statistical Mechanics of Complex Networks, Reviews of Modern Physics, Vol 74 (1), pp. 47 – 97 (2002)
12. Dorogovtsev, S. N., Mendes, J. F. F.: Evolution of Networks, arXiv: cond-mat/0106144v2[cond-mat.stat-mech](2001)
13. Boccaletti, S., Latora, V., Moreno, V., Chavez, M., Hwang, D. – U.: Complex Networks: Structure and Dynamics, DOI:10.1016/j.physrep.2005.10.009 (2006)
14. Zachary, W.: An Information Flow Model for Conflict and Fission in small Groups, Journal of Anthropological Research, Vol 33(4), pp. 452 – 473 (1977)