Cronfa - Swansea University Open Access Repository

This is an author produced version of a paper published in :
CHI'16 CHI Conference on Human Factors in Computing Systems San Jose, CA, USA — May 07 - 12, 2016

Cronfa URL for this paper:
http://cronfa.swan.ac.uk/Record/cronfa26488

Conference contribution :
Lee, A. & Archambault, D. (2016). *Communities Found by Users -- not Algorithms.* CHI'16 CHI Conference on Human Factors in Computing Systems San Jose, CA, USA — May 07 - 12, 2016, (pp. 2396-2400).

http://dx.doi.org/10.1145/2858036.2858071

This article is brought to you by Swansea University. Any person downloading material is agreeing to abide by the terms of the repository licence. Authors are personally responsible for adhering to publisher restrictions or conditions. When uploading content they are required to comply with their publisher agreement and the SHERPA RoMEO database to judge whether or not it is copyright safe to add this version of the paper to this repository.
http://www.swansea.ac.uk/iss/researchsupport/cronfa-support/

Communities Found by Users – not Algorithms: Comparing Human and Algorithmically Generated Communities

Alexandra Lee
Swansea University
741233@swansea.ac.uk

Daniel Archambault
Swansea University
d.w.archambault@swansea.ac.uk

ABSTRACT
Many algorithms have been created to automatically detect community structures in social networks. These algorithms have been studied from the perspective of optimisation extensively. However, which community finding algorithm most closely matches the human notion of communities? In this paper, we conduct a user study to address this question. In our experiment, users collected their own Facebook network and manually annotated it, indicating their social communities. Given this annotation, we run state-of-the-art community finding algorithms on the network and use Normalised Mutual Information (NMI) to compare annotated communities with automatically detected ones. Our results show that the Infomap algorithm has the greatest similarity to user defined communities, with Girvan-Newman and Louvain algorithms also performing well.

Author Keywords
Social Media/Online Communities; Visualisation; Quantitative Methods; Lab Study; Empirical study that tells us about people

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous; Algorithms Experimentation: Human Factors

INTRODUCTION
With the advent of social networks, there is an ever increasing need for the automatic processing of these networks in order discover information contained within them. One of the most basic questions one can ask is what are the social communities present in the network given the relationships between the actors. Many community finding algorithms have been designed and evaluated [9] to address this problem. The focus of these algorithms has been mainly to separate out densely connected areas of the network while breaking as few edges as possible. These algorithms [2, 6, 11, 13, 17, 20] have been created by members of the data mining community and have been evaluated by comparing the results to generated graphs with known ground truth embedded in them [10]. However, these evaluations have not explicitly considered the human-centred perspective of this question.

When evaluating community finding approaches numerically, Normalised Mutual Information (NMI) [5] is often used. NMI can measure the difference between two sets of communities on the same graph. In the evaluations of Lancichinetti and Fortunato [9], NMI was used to compare generated ground truth communities to the communities detected by the algorithm being tested. This metric returns a value in the range $[0, 1]$ with 0 indicating that there was no correspondence between the ground truth and the output of the algorithm and 1 meaning that there is a perfect correspondence between the two. NMI is a useful metric for gauging how closely a community finding algorithm matched a particular desired output.

In this paper, we perform a user study that compares the results of many of the leading state-of-the-art community finding algorithms with human-generated communities. This was carried out to answer the research question "which community finding algorithm most closely matches the human notion of communities?". In our study, twenty participants downloaded and were presented with a node-link visualisation of their egocentric Facebook network. Nodes in the network were labelled with the names of the actors and the participant was asked to manually annotate the social communities present in their network. Both the network structure and annotated social communities were saved anonymously to the machine. Independently, a selection of community finding algorithms that performed well in the study of Lancichinetti and Fortunato [9] were run on the collected graph structure. NMI was used to quantify how closely the detected communities matched the annotated communities for all algorithms.

Results of our analysis indicate that the map equation algorithm (referred to as Infomap) has a significantly higher similarity to user defined communities using NMI. Louvain and Girvan-Newman also have a reasonably high level of similarity with user communities. We also provide evidence that the ranking of Lancichinetti and Fortunato [9] when considering graphs with known truths is similar to the ranking of algorithms when considering user defined communities.

RELATED WORK

Community Finding Algorithm Surveys
Papers have tested community finding algorithms against artificially generated networks. Lancichinetti and Fortunato [9] set out to compare several community finding algorithms against benchmark data sets in order to find algorithms that

most closely match embedded ground truth. This paper provided a rigorous set of tests for community finding algorithms and carried out these tests on these algorithms. The LFR (Lancichinetti-Fortunato-Radicci) [10] benchmark is an algorithm that generates artificial networks that resemble real-world networks with a priori known communities. In Lancichinetti and Fortunato [9], the Infomap algorithm [17, 18] performed the best in their analysis. The algorithms that are tested in our paper were selected from the above article as they were a sample of most highly ranked ones in the survey [9]. The algorithms used are CFinder [13], Girvan-Newman [6], Infomap [17, 18] , Infomod [20], Louvain [2], and OSLOM [11].

Orman et al. [12] modify the LFR algorithm to allow it to produce networks with some typical features of real world networks. Shortcomings of the LFR algorithm that they set out to address are that "the generated networks exhibit a low transitivity and close to zero degree correlation for certain community structures, while according to Newman, real-world networks usually have a clearly non-zero degree correlation, and their transitivity is relatively high" [12]. To find the realism level of generated networks, known topological properties of the generated networks are compared with reference values commonly observed in real-world networks. The researchers find that the proposed modifications to the LFR method lead to more realistic networks in terms of average distance, degree correlation, and centralisation [12].

Aldecoa et al. [1] perform an evaluation of community finding algorithms using complex closed benchmarks and a new type of analysis based on hierarchically clustering the solutions suggested by multiple community detection algorithms [1]. They find that no tested community finding algorithm can consistently give optimal results for all networks and that the new analysis, done by combining multiple algorithms, obtains "quasi-optimal performances in these difficult benchmarks" [1].

This area of related work tests community finding from an optimization perspective while we consider the problem from a human-centred perspective.

User Generated Layouts

A similar methodology to the one employed in our paper has been used to determine representative layouts for network data. In these studies, users are presented with a visualisation or description of a network and are asked to draw what they feel is a representative drawing of the data [14–16]. In one of the first papers on this topic, van Ham and Rogowitz [19] asked users to refine circular and force directed layouts of networks into representative drawings. The study found users tend to clearly define and separate cluster structure in the graph. Dwyer et al. [4] compared user generated layouts and force-directed layouts of networks. In this study, users were asked to arrange a network using both a multi-touch table top display and a mouse and keyboard on a standard desktop. The study found that user-generated layouts outperformed force-directed approaches. In a series of experiments, Purchase et al. [14–16] study how users draw a network given its description and not a visual representation of its structure. One of the main findings of these studies is how the description of the network was communicated (edge list, adjacency list, etc.) greatly influenced the networks drawn.

In the above studies, the properties of user generated layouts were compared to graphs automatically drawn with algorithms. Instead of applying this methodology to graph layout, we employ it to study how users perceive community structures. In our experiment, we supply the participant with a node-link visualisation of their egocentric Facebook network and ask them to annotate their social communities.

Hogan et al. [8] describe an interview-based data collection procedure for social networks. Respondents visually arrange alter names written on pieces of paper during an interview. The paper claims that this procedure allows interviewers to work with respondents to identify the strength of relationships and to efficiently capture ties between alters. Although related, in this paper our focus is not on data collection.

EXPERIMENT

We designed an experiment to study how closely community finding algorithms match the notion of communities from a human-centred perspective. For our experiment, twenty participants were recruited through a post on Facebook. Some participants were expert users and others were more casual users. Data collection and the experiment were carried out in March and April of 2015. We would have liked more participants but the Facebook API change precluded any further data collection.

Social Network Collection with NodeXL

NodeXL [7] is an Excel template that allows the analysis and display of network graphs. Graphs can be imported in a variety of formats. For this experiment the Social Net Importer plugin was used so that Facebook data could be collected. The participant used this application to log into Facebook and gave the application permission to read their mutual friend graph.

Community Annotation Application

D3.js [3] was used to create an interactive application, allowing the user to annotate their communities over their network, through use of a lasso tool. This application was developed by the researcher for the purpose of running the user experiment. The D3 force-directed graph function was used, importing participant's Facebook data to create a visual representation on which the participant could define their perceived communities. A search function was added, as well as buttons to toggle on and off the visibility of node name labels and edges. It was possible for a user to change the node colour by simply clicking on the corresponding node. Communities were defined through use of node colour and the lasso tool that was built into the application.

Experimental Procedure

Twenty participants were recruited for the experiment. Ages varied from 16 to 48, and all were native speakers of English. Participants were given both verbal and written instructions regarding the experimental procedure. Also there was extra

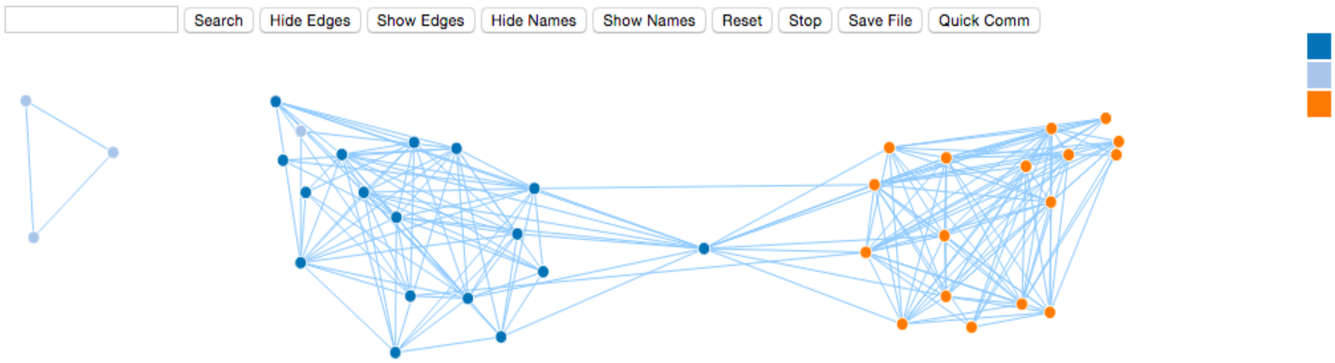

Figure 1. The community annotation application with some nodes coloured and node names hidden.

support available from the researcher if it was required. There were no time limits imposed upon the participants as those with very large friend graphs could take longer to complete the experiment than those with few Facebook friends.

At the beginning of the experiment, the participant was asked to first log in to Facebook using NodeXL. NodeXL presents a selection of data sets relating to the user's Facebook data, but only an edge list of mutual friends was required on this occasion. The participant then saved the acquired data to .csv and .pajek format locally to disk.

The data was then loaded into the D3 application by the participant. The application transformed the acquired Facebook egocentric social network into a node-link representation. Initially, all nodes were dark blue. The participant then sorted these nodes into communities using the simple drag-and-drop interface. The participant could toggle the visibility of edges and name labels as well as change the colour of grouped nodes. When the participant was satisfied with their communities, they were asked to use a selection tool to draw around each community depicting their communities and any community overlap that they felt was necessary. When this was complete, the fully defined community list and graph structure were saved using the participant number assigned to the user. However, when the participant annotated the network all of their friend names were visible on the interface.

The participant data was stored anonymously. In particular, the name fields of each node were deleted and replaced with an identifier (a unique string of 8 letters). This identifier was used to identify the node in the graph, community structure, and overlapping community structure that was identified.

Algorithm Parameters and Network Statistics

All algorithms were run using their default settings, or those suggested by the creators. We did not wish to tweak the parameters as the algorithmically found communities would no longer be simply found by the algorithms and would instead be influenced by human interaction which was not our intent.

Infomap was run with -N 10 parameter; therefore the algorithm assumed that it was an undirected network, partitioned it hierarchically, and output the best result of 10 attempts. Infomod was run with the default 345234 and 10 parameters. 345234 is a random seed and 10 is the number of attempts to partition the network. Oslom was run using the undirected version of the algorithm, along with the mandatory flag -uw (unweighted), with 10 runs. Louvain was run using the NodeXL implementation of the algorithm with no parameters changed from default. Girvan Newman was run using the NodeXL implementation of the algorithm with no parameters changed from default. CFinder was given a maximal allowed time for searching for the cliques of 10 seconds. K-clique size was set to 4 as above this some nodes were not included in the output file.

The largest network had 774 nodes and the smallest had 38. The average networks size was 305 nodes. Only six networks had more than this average but these networks were very large. If we were to remove these largest six networks, the average becomes 163 nodes.

COMMUNITY COMPARISON USING NMI

The community finding algorithms, discussed in related work, were run on the user's egocentric social network. The output of each algorithm was then compared to the user defined communities from the experiment using NMI [5]. Results were given in the range of $0 - 1$, where 0 meant that there was no match between user and algorithmically defined communities and 1 indicated a perfect correlation between algorithmically found and user defined communities.

Algorithmically found and user defined communities were compared using NMI. A Shapiro-Wilk test, with a significance level of $\alpha = 0.05$, was used to determine whether or not the data was normally distributed. It was found that no distribution of algorithm NMI values were normal. As a result a Friedman test, with a significance level of $\alpha = 0.05$, was used. Post-hoc analysis was conducted with the Wilcoxon-Nemenyi-McDonald-Thompson test. In Figure 2, black lines connect pairs of bars with significant differences. The standard error is indicated on each bar. Pairs that showed significant differences in the post-hoc test were Infomap - CFinder,

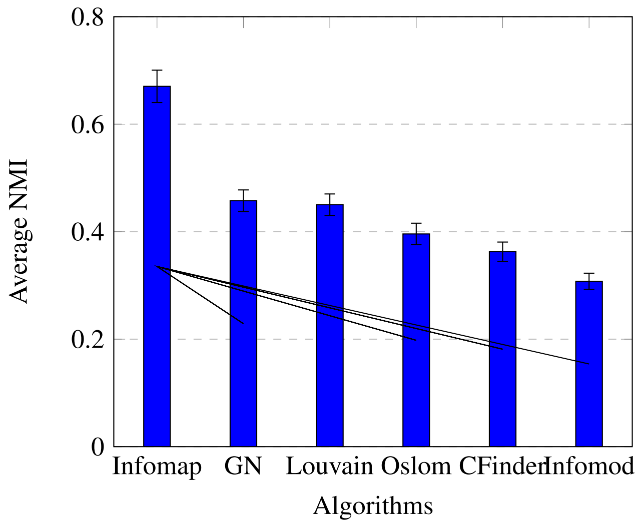

Figure 2. The average NMI scores of all tested algorithms. 5% margin of error is shown by the error bars. Black lines represent pairs of algorithms that were found to have significant differences when a post-hoc test was carried out.

Infomap - Infomod, Infomap - Oslom, and Infomap - Girvan Newman.

DISCUSSION

From the post-hoc test, we can deduce that Infomap is closer to user generated communities (using NMI values) when compared to all other algorithms, except Louvain. Infomap also has the highest average NMI score of all algorithms. Using this finding it is now possible to state that there is evidence that Infomap is the community finding algorithm that bears the most resemblance to user defined communities. Surprisingly Girvan and Newman also performs well, despite it being one of the first community finding algorithms proposed. This would imply that the concept of finding communities using edge betweenness as a measure is still one that has value but also perhaps one that needs to be refined. Oslom and CFinder also perform well, with Infomod having the lowest average NMI score of all algorithms.

It was found in most cases that all other tested algorithms (Louvain, Girvan-Newman, Oslom, CFinder, Infomod) often identified too many communities - the worst case being 36 communities detected for a graph with only 73 nodes. This would suggest that community finding algorithms in general struggle to find communities in graphs with low community structure as there are few heavily connected areas of the social network. It also suggests that, for these cases, community finding algorithms tend to find too many communities when compared to how a user would annotate them based on the names of the actors in the network.

The results of this experiment provide a user-centred perspective on the experiment of Lancichinetti and Fortunato [9] that considers generated graphs with embedded ground truth. In this experiment, users annotate their social communities on top of their own Facebook egocentric networks and we use this as ground truth to compare the results to automated approaches. In both cases, we find that the results are quite similar. In both experiments, Infomap provides the best performance followed by Louvain. Both Girvan-Newman and CFinder perform relatively well. Infomod performs best on simpler graphs with fewer nodes. Thus, its seems that the user's perception of their social communities on Facebook graphs produce very similar results to studies from the perspective of optimisation [9].

CONCLUSION AND FUTURE WORK

In this paper, we present the results of an experiment where we asked users to annotate their own egocentric Facebook network and compared the result to algorithmically generated communities. Algorithms performing well in our experiment are consistent with the top ranking algorithms of Lancichinetti and Fortunato [9], indicating these two studies, taken from very different perspectives, find similar results.

At this time, it would appear that Infomap is the community finding algorithm that is most similar to human defined social communities. This finding is most likely to be due to the non-destructive, flow based method used by the algorithm; implying that users do not consider destroying their social graphs to create communities. That is, users do not remove edges as a way of defining communities but rather leave the underlying social graph structure in place and work around it.

All community finding algorithms had a higher NMI on graphs with an inherent community structure. Considering the findings of Purchase et al. [14–16] for user-generated layouts, a possible limitation of our experiment is the chosen visualisation (node-link) which could have influenced our user-defined communities - nodes were grouped together where there were lots of edges shared between nodes; potentially influencing users into defining communities in areas where there were many grouped nodes. Future work would involve determining if this standard, and easily comprehensible, method for visualising social networks could influence how users annotate their own network.

The human-centred perspective of community finding can offer an important perspective during the development and testing of future community finding algorithms. If an algorithm is frequently able to find communities similar to those a human would expect in an egocentric network with which they have familiarity, this provides some validation, from a human-centred perspective for that algorithm. Although it is unlikely that a user would be contemplating modularity maximisation explicitly whilst defining their social communities, whatever metric they are using should, in some sense, be mirrored by automatic community finding algorithms.

For the purposes of our study (users examining their own network), egocentric networks seemed a logical place to start. Further experiments with different network topologies are needed. In the case of general social networks, a user might not be familiar with all the actors. In the case of a general social network where the user is familiar with all or most actors, we conjecture that higher connectivity and lower modularity would lead to larger communities, but this remains future work.

REFERENCES

1. R. Aldecoa and I. Marín. 2013. Exploring the limits of community detection strategies in complex networks. *Scientific Reports* 3 (2013), 2216 EP –.
2. V. D. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008, 10 (Oct. 2008), P10008. DOI: http://dx.doi.org/10.1088/1742-5468/2008/10/P10008
3. M. Bostock, V. Ogievetsky, and J. Heer. 2011. D3 Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2301–2309. DOI: http://dx.doi.org/10.1109/TVCG.2011.185
4. T. Dwyer, L. Bongshin, D. Fisher, K.I. Quinn, P. Isenberg, G. Robertson, and C. North. 2009. A Comparison of User-Generated and Automatic Graph Layouts. *Visualization and Computer Graphics, IEEE Transactions on* 15, 6 (2009), 961–968.
5. P.A. Estevez, M. Tesmer, C.A. Perez, and J.M. Zurada. 2009. Normalized Mutual Information Feature Selection. *Neural Networks, IEEE Transactions on* 20, 2 (Feb 2009), 189–201. DOI: http://dx.doi.org/10.1109/TNN.2008.2005601
6. M. Girvan and M. E. J. Newman. 2002. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences* 99, 12 (2002), 7821–7826. DOI: http://dx.doi.org/10.1073/pnas.122653799
7. D. Hansen, B. Shneiderman, and M. A. Smith. 2010. *Analyzing Social Media Networks with NodeXL: Insights from a Connected World*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
8. B. Hogan, J. A. Carrasco, and B. Wellman. 2007. Visualizing personal networks: working with participant-aided sociograms. *Field Methods* 19, 2 (2007), 116–144.
9. A. Lancichinetti and S. Fortunato. 2009. Community detection algorithms: A comparative analysis. *Phys. Rev. E* 80 (Nov 2009), 056117. Issue 5. DOI: http://dx.doi.org/10.1103/PhysRevE.80.056117
10. A. Lancichinetti, S. Fortunato, and F. Radicchi. 2008. Benchmark graphs for testing community detection algorithms. *Phys. Rev. E* 78 (Oct 2008), 046110. Issue 4. DOI: http://dx.doi.org/10.1103/PhysRevE.78.046110
11. A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato. 2011. Finding Statistically Significant Communities in Networks. *PLoS ONE* 6, 4 (04 2011), e18961. DOI: http://dx.doi.org/10.1371/journal.pone.0018961
12. G. K. Orman, V. Labatut, and H. Cherifi. 2013. Towards Realistic Artificial Benchmark for Community Detection Algorithms Evaluation. *Int. J. Web Based Communities* 9, 3 (June 2013), 349–370. DOI: http://dx.doi.org/10.1504/IJWBC.2013.054908
13. G. Palla, I. Derenyi, I. Farkas, and T. Vicsek. 2005. Uncovering the overlapping community structure of complex networks in nature and society. *Nature* 435, 7043 (09 06 2005), 814–818. http://dx.doi.org/10.1038/nature03607
14. H.C. Purchase, C. Pilcher, and B. Plimmer. 2012. Graph Drawing Aesthetics – Created by Users, Not Algorithms. *Visualization and Computer Graphics, IEEE Transactions on* 18, 1 (Jan 2012), 81–92. DOI: http://dx.doi.org/10.1109/TVCG.2010.269
15. H. C. Purchase. 2013. Sketched Graph Drawing: A Lesson in Empirical Studies. In *Proc. of Graph Drawing (GD '13)*. LNCS, Vol. 8242. 232–243.
16. H. C. Purchase. 2014. A healthy critical attitude: Revisiting the results of a graph drawing study. *Journal of Graph Algorithms and Applications* 18, 2 (2014), 281–311. DOI: http://dx.doi.org/10.7155/jgaa.00323
17. M. Rosvall, D. Axelsson, and C. T. Bergstrom. 2009. The map equation. *The European Physical Journal Special Topics* 178, 1 (2009), 13–23. DOI: http://dx.doi.org/10.1140/epjst/e2010-01179-1
18. M. Rosvall and C. T. Bergstrom. 2007. An information-theoretic framework for resolving community structure in complex networks. *Proceedings of the National Academy of Sciences* 104, 18 (2007), 7327–7331. DOI: http://dx.doi.org/10.1073/pnas.0611034104
19. F. van Ham and B. Rogowitz. 2008. Perceptual Organization in User-Generated Graph Layouts. *Visualization and Computer Graphics, IEEE Transactions on* 14, 6 (Nov 2008), 1333–1339. DOI: http://dx.doi.org/10.1109/TVCG.2008.155
20. E. Ziv, M. Middendorf, and C. H. Wiggins. 2005. Information-theoretic approach to network modularity. *Physical Review E* 71(4 Pt 2) (Apr 2005), 046117. Epub 2005 Apr 14. http://link.aps.org/abstract/PRE/v71/e046117