Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Are Circles Communities? A Comparative Analysis of Selective Sharing in Google+

Steffen Brauer

Project 1 Report

Fakultät Technik und Informatik
Department Informatik

Faculty of Engineering and Computer Science
Department of Computer Science

Steffen Brauer

Project 1 Report

Eingereicht am: 28. February 2013

Contents

1 Introduction

In Social Media applications like Blogs, Online Social Networks (OSNs), or Micro Feeds, users can communicate and share content in groups. While the audience of direct communication and private messages is explicitly known, content sharing and status messaging are commonly distributed via implicit replication on the platform. A possible large audience of a post is often not considered by the user, but may have significant impact on his personal life. The so called over-sharing of content can be avoided using features for selective sharing [1], which are provided by several Social Media services. The concept of selective sharing was designed to support a context-dependent publication behaviour. In different contexts like work, family, or friends, a person can act differently according to appropriate norms and accepted conventions. These so called facets [2] or foci [3] describe different social aspects in the life of a person and provide the theoretical background of selective sharing features.

Released only in 2011, Google+ is a latecomer in the world of Online Social Networks. One of its characteristic features are the circles. Circles allow users to group their contacts on the platform in different communities, and manage their different social facets. Preconfigured circles are friends, family and acquaintances, as well as all following users. It is also possible to rename the default circles or add customized ones. The range of users, who can read a post, is limited by the members of the target circle that was selected when the post was created. Besides selective sharing, the Google+ circles also enable selective reception of content streams generated by the contacts in each circle. While the circles are private to the user per default, there is the option to share circles. Another interesting feature of Google+ is directed edges, which are created by adding users to circles. Outgoing edges are named 'In your circles' and incoming edges 'In circles of others'. These directed links in Google+ make it a hybrid network between typical OSNs like Facebook and Micro-blogging services like Twitter [4].

We are interested in understanding the group structures within OSNs [5]. In this work, we investigate the characteristics of circles within the Google+ graph and compare to well-known group structures like communities. We want to find out how particular these circles are, and try to identify a clear signature of their structure if available.

A circle is actively created and shaped by its owner, possibly based on a personal perspective. In contrast, communities or groups follow a subscribe model and are open for joining by different users who share a common attribute. Given such difference in composition, it is natural to question the similarity of the results. Using a variety of empirical data sets, we explore the following details:

1. Are circles pronounced structures in the social graph, well distinguished from the underlying network?
2. Are circles like communities or other group structures of classical social networks, or do circles show other, dedicated characteristics?

To answer these question, we analyzed the data sets using different scoring functions. We base our results on four scoring functions from the field of community detection that focus on specific characteristics of network communities. Our findings indicate that circles do admit a very pronounced structure that differs from other groups by its connectivity with the remaining social graph. While traditional communities are rather closed groups with few relations to the outside, circles in Google+ are densely connected internally *and* externally. This makes them appear like communities with many additional transit links.

The remainder of the paper is structured as follows. We continue with reviewing of related work on Google+ and its circle feature. In Section 3, we discuss the theoretical differences between circles and traditional communities. The following Section 4 characterizes the Google+ and LiveJournal data sets we use in our evaluation. In Section 5, we evaluate and discuss the questions stated above. We conclude in Section 6 and give an outlook.

2 Related Work

Since its release, Google+ has attracted research. Schiöberg et al. [4] traced the social graph in the beta and initial public phase. They found that the asymmetric relations in Google+ lead to a hybrid form of classic social network like Facebook and Social Media services like Twitter.

Magno et al. [6] also crawled Google+ in its creation phase. While they evaluate typical metrics of Social Network Analysis, they also include reciprocity. Their Relation Reciprocity describes the correlation between in- and out-degrees for a given vertex or set of vertices.

Besides the typical social network features, circles are an interesting field of social network analysis. Kairam et al. [1] focus on the selective sharing aspect and how it is used in Google+. They apply the theory of facets of a person's life. Each facet covers another group of people someone interacts with in the daily life. While some people share the same information with all people they know, some people want to share special information just with some people. Kairam et al. show in their work, that circles on Goolge+ correspond to these natural groupings.

McCauley and Leskovec [7] propose a model for automatically discovering circles in a given ego-network. Based on the ego-network of a user, they formulate the circle detection as a clustering problem on the users' ego-network. The authors aim at modelling properties with respect to the assumed characteristics of circles: i) vertices in a circle share a common property or aspect, ii) each circle is formed by a different aspect and iii) circles can overlap that means that strong circles can be within weaker ones. These characteristics define circles as a kind of network community and we want to evaluate if these characteristics form pronounced structures in the social graph.

Fang et al.[8] investigate the impact of the circle-sharing feature for the network growth and structure of the Google+ social network. According to their clustering, shared circles can be categorized into two main groups: communities, which have a high link density and reciprocity as well as celebrities, which have a low in-circle density, low reciprocity but a very high in-degree. These two categories show the main motivation of users to share their circles: they want to share group of users, which may have a common attribute or represent a facet of their life, or they want to share a group of popular people.

Yang and Leskovec [9] evaluate community metrics on data sets, which include ground-truth communities. The authors map ground-truth communities to crawled network communities, which are explicitly labelled in the provided evaluation data. They use 230 different social, collaboration and information networks to test 13 scoring functions, which characterize how well a set of vertices is connected. One of their findings is, that the scoring functions correlate and can be grouped in four subsets based on the community characteristics they measure. We base our work on their categorization of scoring functions and use two data sets which are provided by the authors.

3 Circles vs. Communities

Circles are one of the main features in Google+. While other social networks like Facebook support features for selective sharing, Google+ forces the user to put new contacts in circles. While some users try to avoid this categorization by putting all contacts in one circle, the majority actively uses this feature [10]. The mapping from contacts to users is only visible to the creator of the circle, which makes quantitative evaluations on circles difficult from the public perspective. The only way to extract circles from Google+ without asking users to manually publish their circles is to search for posts with shared circles. Users are able to share circles they created by others. While reasons for sharing a circle are manifold, this feature creates an opportunity of accessing the structure of circles. The limitation of shared circles is that they may be created only for sharing and do not represent a actual facet of the creator. Fang et al. [8] found that there are two main categories of shared-circles. Circles, that cover communities. They have a very high density and a high reciprocity with the circle owner. And there are circles which cover very popular users. The popularity is determined by the users in-degree.

Communities can be found in different types of network like social, biological or information networks. They all refer to a common attribute of the vertices, which stimulates connections. A general definition of communities in networks is that a community is a set of vertices, which has many connections within the set and that there are just view connections from the community to the remaining network [11, 9]. Like classical communities, circles are created around a common attribute, which may only be reasonable for the creator of the circle and is not strong enough to form a community-like structure. Another limitation of this creation process is that only vertices with the common attribute can be added to the circle, which are in the ego-network of the creator.

4 Data Description

In the remainder of the paper we refer to the social graph as the directed graph $G(V, E)$, where V are the user profiles and E the relations between them. A vertex $v \in V$ has an ID to identify it. Other available attributes are not considered, because we only want to evaluate the structural properties of the social graph. Based on the directed Graph $G(V, E)$, we define n as the number of vertices in the graph $n = |V|$ and m as the number of edges $m = |E|$. The scoring function f is called for a circle C with n_C as the number of vertices in C $n_C = |C|$, m_C as the number of edges in C and c_C as the number of edges on the boundary of C. $d(v)$ is the degree of a vertex v.

Table 1: Nomenclature

Notation	Description
n	Number of vertices in the graph
m	Number of edges in the graph
f	Scoring function
C	Circle/ community
n_C	Number of vertices in C
m_C	Number of edges in C
c_C	Number of edges at the boundary of C
$d(v)$	Degree of vertex v

4.1 Google+

Even if the contacts of a user are publicly available and there are shared circles, it is difficult to load the data from Google+, because the official API provides no call for loading relationships. Schiöberg et al. [4] use the requests the Google+ Web interface sends to get information from the Google+ servers. These calls can be made visible by using a debug tool for web browser. By changing the G+ ID in the request and a key word for in- or outgoing relations, the Google+ graph can be extracted. Because it is very time consuming to implement such a social network crawler, we use a the only available data set provided by McCauley and Leskovec [7], which include circles. It covers 133 ego-networks of users, who share at least two circles. The authors develop an algorithm, which is able to find circles in ego-networks of users in different social networks. While they also use Facebook and Twitter for their evaluations, they crawl a Google+ data set. Since circles are per default private to users, they chose user, who share at least two circles and crawl their ego-network. The ego-network of users covers all vertices, they are connected to and all edges between these vertices. By using this method, they collect the ego-networks from 132 Google+ users including 468 circles. Even if the data set only consists of ego-networks, by joining all ego-networks a large connected component if formed with 107.614 vertices and 13.673.453 edges.

Data, which is generated with this method, is biased towards too dense components, which are connected by a few edges between each other. Figure 1 shows the scheme of the data set. It consists of ego-networks (light blue areas) with the owner (red vertex) in it. The owner has connections to each other blue vertex in the ego-network. Because

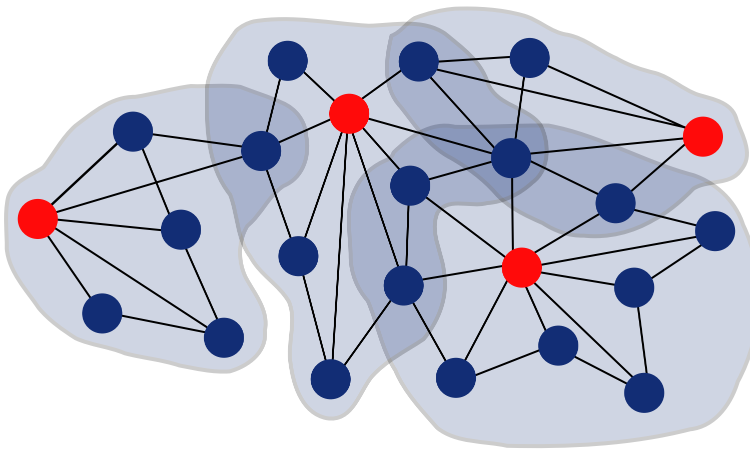


Figure 1: Scheme of the data set from McAuley and Leskovec [7]: Red vertices are the owner of the ego-network, the vertices in the overlapping ego-networks are the bridges between the ego-networks.

several vertices appear in more than one ego-network, the joint graph of all ego-networks is a whole connected component. We find that 93,5 % of the ego-network in the data set overlap. This means that they share at least one common vertex. The overlap of ego-networks is shown in Figure 2. More than 55.000 vertices are only in one and around 14.000 in two networks. Comparing to this high count of vertices, there just a few vertices, which are members in more than 50 ego-networks. These vertices have a high impact on the connectivity of the data set.

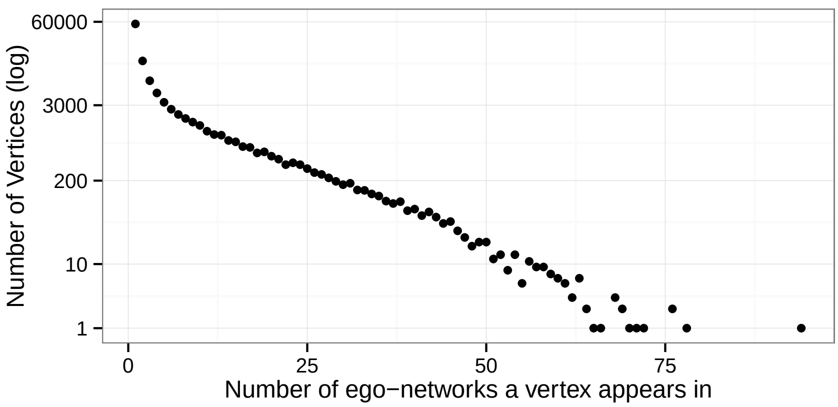


Figure 2: Membership of vertices in ego-networks. Number of Vertices is in log scale

Before we start with the evaluation, we characterize the data set by investigating common structural features of Online Social Networks: Degree Distribution, Clustering Coefficient and the Node Separation [12].

4.1.1 Degree Distribution

The degree $d(v)$ of a vertex v in a graph G is defined as the number of connected edges. Previous research on Online Social Networks found that the distribution of the degree often follows a power-law distribution. Magno et al. [6] even found a power-law distribution in their Google+ data set. To investigate the degree distribution in the McAuley and Leskovec [7] dataset, we use the method introduced by Clauset et al. [13]. The authors stress that determining a power-law distribution by using only plots in not satisfying. By following the proposed method, we cannot find a power-law distribution in the degree distribution of the provided data set. But we find a fit for a log-normal distribution of the in-degree of the vertices (Figure 3).

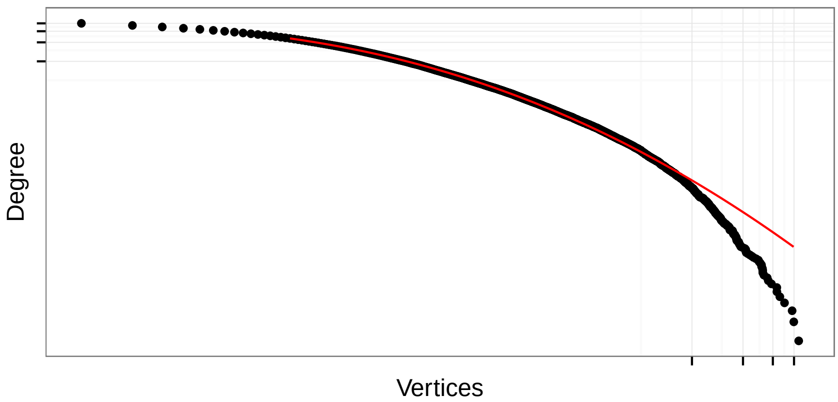


Figure 3: In-degree Distribution of the Google+ data set. Red line shows the fit of the log-normal distribution

4.1.2 Clustering Coefficient

Besides the degree distribution, the Clustering Coefficient (CC) is a common measure on OSN. It is based on the local clustering structure of vertices. The neighbours of a vertex tend to connect to each other as well [12]. The phenomenon is formalized by the number of triangles a vertex is part of in relation the maximum number of triangles a vertex could be in [6]. The average clustering coefficient is 0.4901 on the used data set. In comparison to the measurements of Magno et al. we observe a general lower coefficient. Gong et al. [14] evaluate the CC in the creation phase of Google+. The highest value they observe at the very beginning of Google+ is about 0.32, which is much smaller than our observed values.

4.1.3 Node Separation

As a third characteristic of OSN we measure the Node Separation. This characteristic is one of the early objectives in social network research, known as the small world problem, published by Milgram 1967 [15]. He discovered that all people in the US know each other via 6 other people. To make statements about the Node Separation in an OSN, two metrics are common: The network diameter and the average path length. The diameter of a network is the length of the longest shortest path between all vertices. While we observe a diameter of 6, Gong et al [14] observe a value of 6 on their data set and Magno et al. [6] about 19. We also evaluate the average path length of the dataset with 3.32 (Magno et al. 5.9).

Comparing the statistics of our evaluation with the evaluation performed by Magno et al. [6], it shows that the data set we use is significantly smaller and more connected. Table 2 shows an overview of the two data sets. Magno et al. crawl a data set that has 326 times more vertices than our data set but there is a lot smaller difference in the edge count. The Magno data set only has 42 times more edges than ours. The higher connectivity of the data set in our evaluation is also followed from the average degree of vertices in the networks. While the Magno network has an average of 16 edges per vertex, the vertices in our data set have an average degree of 127. This difference is caused by the crawling methods, while Magno et al uses a Breath-First-Search, the McAuley and Leskovec data set is crawled around defined users and their ego-networks. It is important to consider the high connectivity of our data set in the evaluation, because we investigate the structure of circles within the network, focusing on their intern connectivity. Metrics applied to the circles, which are based on the vertex edge ratio will produce high values, because of the general high density in the network, compared to the result of the same metric on other Google+ data sets.

4.2 Other Data Sets

Besides the Google+ data set described above, we use several other social network graphs to verify our evaluation. An overview is given in Table 3. The Twitter data

Table 2: Statistical comparison of the McAuley and Leskovec [7] and Magno et al.[6] data sets. (ASP: Average Shortest Path)

Metric	**Magno et al.**	**McAuley and Leskovec**
Vertices	35.114.957	107.614
Edges	575.141.097	13.673.453
Diameter	19	13
ASP	5.9	3.32
Degree		
distribution (in)	power-law $\alpha = 1.3$	log-normal
distribution (out)	power-law $\alpha = 1.2$	-
average degree (in)	16.4	127
average degree (out)	16.4	189

Table 3: Comparison of the evaluated Data Sets

Graphs	**Google+ [7]**	**Twitter [7]**	**LiveJournal [9]**	**Orkut [16]**
Vertices	107,614	81,306	3,997,962	3,072,441
Edges	13,673,453	1,768,149	34,681,189	117,185,083
Type	directed	directed	undirected	undirected
Structure	Circles	Circles	Communities	Communities
# Communities	468	100	5000	5000

set is also provided by McAuley and Leskovec [7] and the included communities are created by the same feature as Google+ circles. To compare these circles with classical communities, we also use data sets of LiveJournal [9] and Orkut [16]. We use the top 5000 communities of these two graphs ranked by their size, to save computational resources. The main differences between the two data sets is their edge type. While Google+ and Twitter use directed edges to represent a link between its users, LiveJournal and Orkut use undirected connections. This can have an impact on our evaluation, because a bidirectional relation between two vertices is represented in undirected Graphs by on the edge and in directed by two.

5 Evaluation

The goal of our evaluation is to investigate if the characteristics of classic network communities fit to circles in Google+. Here is the first question if circles form remarkable structures in the social graph. The second question is how these structures act like classic communities in typical social networks.

To quantify the definition of communities, we use scoring functions from the field of community detection. They score set of vertices with a high value, if they match the idea of community, the function is able to identify.

Internal Connectivity: Average Degree A communities should consist of a set of very highly connection vertices. A scoring function which focus on the internal connectivity is the Average Degree [17]. In the definition provided by Yang and Leskovec [9]

$$f(C) = \frac{2m_C}{n_C}, \tag{1}$$

the degree of the vertices in C is calculated by the double of the edges within the community $2m_C$ to get the links on each vertex, divided by the number of vertices n_C in C. This function produces values depending on the actual size and density of the studied network.

External Connectivity: Ratio Cut The second characteristic of a community is that it is separated from the rest of the network. That means, the vertices in the set form just a few connections to the remaining network. The Ratio Cut function is based on the edges which are at the border of a community c_C and divided them by the product of the number of vertices within C and the complement of C[11].

$$f(C) = \frac{c_C}{n_C(n - n_C)} \tag{2}$$

By using the vertices as the characteristic of the community, the connectivity within the community is not mentioned at all.

Combined Internal and External Connectivity: Conductance While Average Degree and Ratio Cut only focus on one aspect of the relation between the community and network, the Conductance metric covers both aspects. This is achieved by measuring the total edges pointing outside the community in relation to the total degree of the vertices within the community [18].

$$f(C) = \frac{c_C}{2m_C + c_C} \tag{3}$$

The Conductance scoring function is able to capture our loose definition of a community. A good community will have a low score, because there should be just a few edges in comparison to the edges within the community.

Modularity A very popular scoring function is Modularity introduced by Newman[19]. It uses a so called null model to verify if the given network has a community structure or not [11]. The model used in our work, proposed by Newman and Girvon [20] consists of a randomized graph, which has the same degree sequence as the original graph. Based on the null model, the Modularity is positive, if the number of edges inside the community is higher than the number of edges in the null model.

$$f(C) = \frac{1}{2m}(m_C - E(m_C)) \tag{4}$$

5.1 Circle Structures

The first question of our evaluation is, whether circles form pronounced structures in the social graph, which are comparable to the common understanding of network communities. To answer this question, we apply the four scoring functions mentioned above to the circles available in the Google+ data set and to randomly selected sets with the same size as the circles from the graph. If the circles form pronounced community-like structures, the functions would score the circles with better values than die randomly selected sets of vertices.

Figure 4 shows the ranked scores for each circle or randomly generated set. All four functions perform better on the circles than on the random sets. All four functions show an exponential distribution of their scores. This shows, that there is always a tail of circles, which have a poor community structure. A reason for the tail of this distribution can be found in the work of Fang et al. [8], they found that there are two main categories of shared-circles in Google+. Communities, which are highly connected and celebrities, which have a low connectivity but the members are very popular. Their first category of community circles produce high scores and a circle of e.g. pop singers have a low intern connectivity and are scored with a low score.

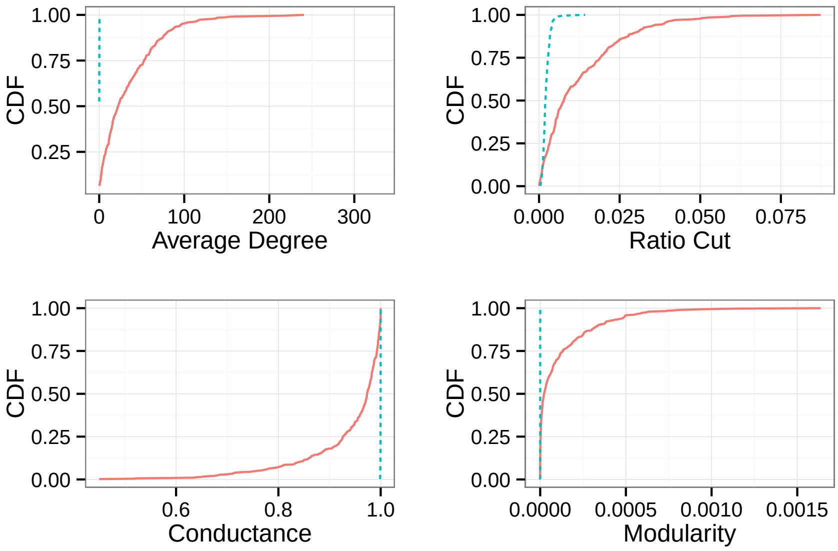
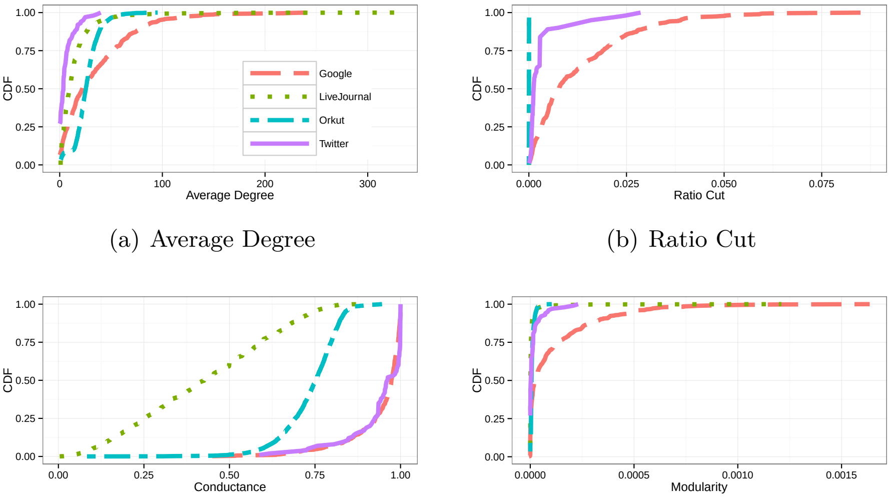


Figure 4: Group Metrics applied to Circles in the dataset (red solid) and group metrics applied to sets of randomly chosen vertices (blue dashed).

While the circle scores of Average Degree, Ratio Cut and Conductance converge at the end of the tail with the random sets, Modularity converges already in the middle of the distribution. The Average Degree is widely distributed from 250 to nearly 0, on the contrary the random sets all produce scores around 0. This is caused by the random selection, which selects vertices, which may not be directly connected to each other. This lead to a score of 0 if there are no edges between the vertices in the random set. Looking at the result of Ratio Cut, there is a small peak for the randomly selected vertices but the main part of the values is constantly over 0 with a decrease at the end. The effect of a decreasing tail is also visible for the circle scores. The Conductance function shows on the random sets a constant score around 1. The reason lies in the low number of edges in the random set and the large number of edges on the boundary. For the circle scores, the Conductance function shows positive scores for a very high rate of circles. Summing up the results of the first question, we show that the scoring functions, which score set of vertices by community characteristics, produce much higher values for the circles in the Google+ data set than on randomly selected set of vertices. So we show that circles form pronounced structures which are differ from the entire graph.



(a) Average Degree (b) Ratio Cut

(c) Conductance (d) Modularity

Figure 5: CDF for the Scores of the scoring function applied to the communities in the four data sets.

5.2 Circles vs. classical Communities

Our second question was about whether circles have a community like structures and are they similar distinctive as communities, which are observed in classic social networks. We use the data sets from LiveJournal and Orkut networks (Section 4.2) to compare the scores of the four scoring functions on circle structures (Google+ and Twitter) and classical communities.

Figure 5(a) shows the CDF of the four social network graphs for the Average Degree function. While there is a large difference in the maximum score in the data sets (Twitter: 40, LiveJournal: 332), there is no significant difference in the distribution of the community types. Google+ shows the lowest Average Degree.

The Ratio Cut function measures the external connectivity of a given set of vertices. If the score is low, there a just some connections to the remaining network, which is a characteristic of a well connected community. The values for the two networks with circles are mostly higher (Twitter has a mean of 6 and Google+ 34) than for the classical communities, which are all around 0 (Figure 5(b)). This shows, that circles are mostly not very well separated from the rest of the network. A reason for this phenomenon lies in the perspective of the creators, who may not be connected with the whole community, they want to share. This result is supported by Fang et al.[8], who found that sharing a circle leads to a densification of community circles, because missing members of the community can create connections to user, they did not connect yet.

The Conductance function, which scores internally dense and externally few connected communities with a low value, shows the biggest difference for the data sets. While Twitter and Google+ produce mostly very low scores, the other two data sets behave differently. LiveJournal shows a nearly linear distribution and Orkut lies in the middle of the two previously described curves. The Conductance values are generally lower for the LiveJournal and Orkut data, but there are a few circles, which produce better scores than the best communities.

The CDFs for the Modulartiy scoring function, plotted in Figure 5(d), all show a very high increase. While Twitter, LiveJournal and Orkut nearly overlap, Google+ shows a smoother curve and a higher maximum value.

Summing up the results of this evaluation question, we can see, that circles internally show the same structure as classical communities. But they are not as much separated from the rest of the network. Besides the edge structure inside the circles, a possible reason for this could lie in the directed edges of the Twitter and Google+ data set. But this is improbable because we did not distinguish between the two edge types in the scoring functions.

6 Conclusion

By introducing the concept of circles to Google+, users can transfer the different facets of their social live to the management of their online contacts. Based on a Google+ data set with shared circles we showed a characterization the Google+ social graph. Using two social networks with circle structures and two with classic communities, we show that i) circles form pronounced community like structures in Google+ and that ii) circles are not as good separated from the rest of the network as classical communities.

In our future work, we try to reproduce this result in another data set, which represents the global characteristics of Google+ better than the one used in this work. Another open question is how the structure of circles differ by switching the perspective from a global to an ego-centred view. While only public shared circles were available in this evaluation, the private circles would provide a deeper insight into the user orientated view of circles.

References

[1] S. Kairam, M. Brzozowski, D. Huffaker, and E. Chi, “Talking in circles: Selective sharing in google+,” in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems.* ACM, 2012, pp. 1065–1074.

[2] S. D. Farnham and E. F. Churchill, “Faceted identity, faceted lives: social and technical issues with being yourself online,” in *Proceedings of the ACM 2011 conference on Computer supported cooperative work*, ser. CSCW ’11. New York, NY, USA: ACM, 2011, pp. 359–368.

[3] S. L. Feld, “The focused organization of social ties,” *American journal of sociology*, pp. 1015–1035, 1981.

[4] D. Schiöberg, S. Schmid, F. Schneider, S. Uhlig, H. Schiöberg, and A. Feldmann, “Tracing the birth of an osn: Social graph and profile analysis in google+,” in *Proceedings of the 3rd Annual ACM Web Science Conference.* ACM, 2012, pp. 265–274.

[5] S. Brauer and T. C. Schmidt, “Group Formation in eLearning-enabled Online Social Networks,” in *Proc. of the International Conference Interactive Computer aided Learning (ICL’12)*, M. E. Auer, Ed. Piscataway, NJ, USA: IEEEPress, Sep. 2012.

[6] G. Magno, G. Comarela, D. Saez-Trumper, M. Cha, and V. Almeida, “New kid on the block: exploring the google+ social graph,” in *Proceedings of the 2012 ACM conference on Internet measurement conference*, ser. IMC ’12. New York, NY, USA: ACM, 2012, pp. 159–170.

[7] J. McAuley and J. Leskovec, “Learning to discover social circles in ego networks,” in *Advances in Neural Information Processing Systems 25*, 2012, pp. 548–556.

[8] L. Fang, A. Fabrikant, and K. LeFevre, “Look Who I Found: Understanding the Effects of Sharing Curated Friend Groups,” in *Proceedings of the 3rd Annual ACM Web Science Conference*, ser. WebSci ’12. New York, NY, USA: ACM, 2012, pp. 95–104.

[9] J. Yang and J. Leskovec, “Defining and Evaluating Network Communities based on Ground-truth,” in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, ser. MDS ’12. New York, NY, USA: ACM, 2012, pp. 3:1–3:8.

[10] J. Watson, A. Besmer, and H. R. Lipford, "+Your Circles: Sharing Behavior on Google+," in *Proceedings of the Eighth Symposium on Usable Privacy and Security*, ser. SOUPS '12. New York, NY, USA: ACM, 2012, pp. 12:1–12:9.

[11] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, no. 3-5, pp. 75–174, 2010.

[12] A. Sala, L. Cao, C. Wilson, R. Zablit, H. Zheng, and B. Y. Zhao, "Measurement-calibrated graph models for social network experiments," in *Proceedings of the 19th international conference on World wide web*, ser. WWW '10. New York, NY, USA: ACM, 2010, pp. 861–870.

[13] A. Clauset, C. R. Shalizi, and M. E. Newman, "Power-law distributions in empirical data," *SIAM review*, vol. 51, no. 4, pp. 661–703, 2009.

[14] N. Z. Gong, W. Xu, L. Huang, P. Mittal, E. Stefanov, V. Sekar, and D. Song, "Evolution of Social-Attribute Networks: Measurements, Modeling, and Implications using Google+," in *Proceedings of the 2012 ACM conference on Internet measurement conference.* ACM, 2012, pp. 131–144.

[15] S. Milgram, "The small world problem," *Psychology today*, vol. 2, no. 1, pp. 60–67, 1967.

[16] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee, "Measurement and Analysis of Online Social Networks," in *Proceedings of the 5th ACM/Usenix Internet Measurement Conference (IMC'07)*, San Diego, CA, October 2007.

[17] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *PNAS*, vol. 101, no. 9, 2004.

[18] J. Shi and J. Malik, "Normalized cuts and image segmentation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 22, no. 8, pp. 888–905, 2000.

[19] M. E. Newman, "Modularity and community structure in networks," *Proceedings of the National Academy of Sciences*, vol. 103, no. 23, pp. 8577–8582, 2006.

[20] M. E. Newman and M. Girvan, "Finding and evaluating community structure in networks," *Physical review E*, vol. 69, no. 2, p. 026113, 2004.