# Unfolding the Core Structure of the Reciprocal Graph of a Massive Online Social Network

Braulio Dumba and Zhi-Li Zhang

University of Minnesota, Twin Cities, MN, USA,
{braulio,zhzhang}@cs.umn.edu

**Abstract.** Google+ (G+ in short) is a *directed* online social network where nodes have either *reciprocal* (bidirectional) edges or *parasocial* (one-way) edges. As reciprocal edges represent strong social ties, we study the core structure of the subgraph formed by them, referred to as the *reciprocal network* of G+. We develop an effective three-step procedure to *hierarchically* extract and unfold the *core* structure of this reciprocal network. This procedure builds up and generalizes ideas from the existing k-shell decomposition and clique percolation approaches, and produces higher-level representations of the core structure of the G+ reciprocal network. Our analysis shows that there are seven subgraphs ("communities") comprising of dense clusters of cliques lying at the center of the core structure of the G+ reciprocal network, through which other communities of cliques are richly connected. Together they form the core to which "peripheral" sparse subgraphs are attached.

**Keywords:** Reciprocal Network ·Google+ ·Network Core ·Reciprocity

## 1 Introduction

Many online social networks (OSNs) such as Twitter, Google+, Flickr contain both *reciprocal* edges, i.e., edges that have already been linked back, and *parasocial* edges, i.e., edges that have not been or is not linked back [1], and thus directed in nature. Reciprocity is defined as the ratio of the number of reciprocal edges to the total number of edges in the network, and has been widely studied in the literature in various contexts, see, e.g., [1–6]. It is believed to reciprocity plays an important role in the structural properties, formation and evolution of online social networks. Empirical studies have shown that many OSNs exhibit a nontrivial amount of reciprocity: Twitter is estimated to have a reciprocity value of 0.22 [7], Google+ 0.32 [8] and Flickr 0.62 [9].

Reciprocal edges represent the most stable type of connections or relations in directed OSNs: for example, in Twitter it represents users are mutually "following" each other, and in Google+ it represents two users are in each other's circles. Hence, reciprocal edges reflect strong ties between nodes or users [10–12]. Most existing studies have focused on reciprocity (a single-valued aggregate metric) to characterize massive *directed* OSNs, which we believe is inadequate. Instead, we consider the *reciprocal graph* (or *reciprocal network*) of a directed

OSN – namely, the *bidirectional* subgraph formed by the reciprocal edges among users in a directed OSN (see Fig. 1 for an illustration). In a sense, this reciprocal network can be viewed as the stable "skeleton" network of the directed OSN that holds it together. We are interested in analyzing and uncovering the *core* structural properties of the reciprocal network of a directed OSN, as they could reveal the possible organizing principles shaping the observed network topology of an OSN [2].

Using Google+ (thereafter referred to as *G+* in short) as a case study, in this paper we perform a comprehensive empirical analysis of the "core structure" of the reciprocal network of G+. Based on a massive G+ dataset (see Sect. 2 for a brief overview of G+ and a description of the dataset), we find that out of more than 74 million nodes and $\approx 1.4$ billion edges in (a snapshot of) the directed G+ OSN, more than two-third of the nodes are part of G+'s *reciprocal* network and more than a third of the edges are reciprocal edges (with a reciprocity value of roughly 0.34). This reciprocal network contains a *giant connected subgraph* with more than 40 million nodes and close to 400 million edges (see Sect. 3 for more details). Existence of this massive (giant connected) reciprocal (sub)graph in G+ raises many interesting and challenging questions. How is this reciprocal network formed? Does it contain a "core" network structure? If yes, what does this structure look like?

In an attempt to address these questions, we develop an effective three-step procedure to *hierarchically* extract and unfold the *core* structure of G+'s reciprocal network, building up and generalizing ideas from the existing k-shell decomposition and clique percolation approaches. i) We first apply (a modified version of) the k-shell decomposition method to prune nodes and edges of sparse subgraphs that are likely to lie at the periphery of the G+ reciprocal network (see Sect. 4). The standard k-shell decomposition method has been proposed to extract the "core" of a network, e.g., that of the Internet AS graph [13]. However, directly applying this method to the G+ reciprocal yields a final graph – a clique of 298 nodes (the maximum clique of the G+ reciprocal network) that consists of a close-knit community of users in Taiwan – which is unlikely to lie at the "core" of the G+ reciprocal network (see discussion in Sect. 6, where we show this clique in fact lies more at the outer ring of G+'s dense core structure). Instead, we stop the k-shell decomposition when the giant connected component breaks down into two (or more) pieces, each containing a dense subgraph (e.g., a large clique). This process yields a dense "core" subgraph of the G+ reciprocal network with approximately 50K nodes and 7M edges. ii) Given this dense "core" subgraph, we first compute the maximal clique that each node is part of (using a simplified Bron-Kerbosh algorithm), and then form a new *directed* (hyber)graph – a form of clique percolation [14], where the vertices are (unique) cliques of various sizes, and there exists a directed edge from clique $C_i$ to clique $C_j$ if half of the nodes in $C_i$ are contained in $C_j$ (see Sect. 5). This new (hyber)graph provides a higher-level representation of the dense core graph of the G+ reciprocal network: the intuition is that the maximal clique containing each node $v$ represents the most stable structure that node $v$ is part of, and the directed edge in a sense reflects the “attraction” (or “gravitational pull”) that one clique (constellation) has over the other. We find that this (hyper)graph of cliques comprises of 2000+ connected components (CCs). iii) Finally, considering these CCs as the core “community” structures (a dense cluster of cliques) of the G+ reciprocal network, we define three metrics to study the relations among these CCs in the underlying G+ reciprocal network: the number of nodes shared by two CCs, the number of nodes that are neighbors in the two CCs, and the number of edges connecting these neighboring nodes (see Sect. 6). These metrics produce a set of new (hyber)graphs that succinctly summarize the (high-level) structural relations among the core “community” structures and provide a “big picture” view of the core structure of the G+ reciprocal network and how it is formed. In particular, we find that there are seven CCs that lie at the center of this core structure through which the other CCs are most richly connected. In Sect. 7, we conclude the paper with a brief discussion of the future work.

We summarize the major contributions of our paper as follows. To the best our knowledge, our paper is the first study on the core structure of a “reciprocal network” extracted from a massive *directed* social graph. While this paper focuses on G+, we believe that our approach is applicable to other directed OSNs.

- We develop an effective three-step procedure to *hierarchically* extract and unfold the *core* structure of a reciprocal network arising from a directed OSN.
- We apply our method to the reciprocal network of the massive Google+ social network, and unfold its core structure. In particular, we find that there are seven subgraphs (“communities”) comprising of dense clusters of cliques that lie at the center of the core structure of the G+ reciprocal networks, through which other communities of cliques are richly connected; together they form the core to which other nodes and edges that are part of sparse subgraphs on the peripherals of the network are attached.

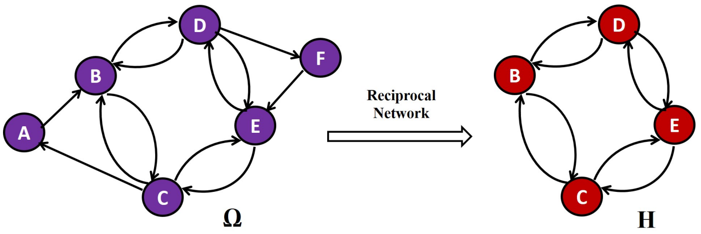


**Fig. 1.** Illustration of the reciprocal network ($H$) of a directed graph ($\Omega$). Specifically, $(B, C)$, $(C, B)$, $(B, D)$, $(D, B)$, $(D, E)$, $(E, D)$, $(C, E)$, $(E, C)$ are reciprocal edges; $(A, B)$, $(C, A)$, $(D, F)$, $(F, E)$ are parasocial edges. The reciprocity of $\Omega$ is $8/12 = 0.67$

## 2 Google+ Overview and Dataset

In this section, we briefly describe key features of the Google+ service and a summary of our dataset.

**Platform Description**: On June 2011 Google launched its own social networking service called Google+ (G+). The platform was announced as a new generation of social network. Previous works in the literature [5, 6] claim that G+ cannot be classified as particularly asymmetric (Twitter-like), but it is also not as symmetric (Facebook-like) because G+ features have some similarity to both Facebook and Twitter. Therefore, they labelled G+ as a hybrid online social network[5]. Similar to Twitter (and different from Facebook) the relationships in G+ are unidirectional. In graph-theoretical terms, if user[1] x follows user y this relationship can be represented as a directed social edge (x, y); if user y also has a directed social edge (y,x), the relationship x, y is called symmetric[15]. Similar to Facebook, each user has a stream, where any activity performed by the user appears (like the Facebook wall). For more information about the features of G+ the reader is referred to [16, 17].

**Dataset**: We obtained our dataset from an earlier study on G+ [6]. The dataset is a directed graph (denoted as $\Gamma$) of the social links of the users[2] in $G$+, collected from August 24th, 2012 to September 10th, 2012. It consists of 74,419,981 nodes and 1,396,943,404 edges. We use Breadth-First-Search (BFS) to extract the *largest weakly connected component* (LWCC) of $\Gamma$. We label the extracted LWCC as subgraph $\Omega$. Since the users $\Omega$ form the most important component of the G+ network [6], we extract the *reciprocal network* of G+ from the $\Omega$ subgraph (see Sect. 3). The main characteristics of $\Gamma$ and $\Omega$ are summarized in the left part of Table 1, where density is defined as $|E|/[|V|(|V|-1)$ for a *directed* graph, and $2|E|/[|V|(|V|-1)$ for an *undirected* graph – here $|V|$ is the number of nodes and $|E|$ is the number of edge.

**Table 1.** Main characteristics of G+ dataset: $\Gamma$ - original G+ network; $\Omega$ - extracted largest weakly connected component of $\Gamma$; $H$ - extracted reciprocal network of G+

| | $\Gamma$ | $\Omega$ | $H$ |
|---|---|---|---|
| # nodes | $74,419,981$ | $66,237,724$ | $40,403,216$ |
| # edges | $1,396,943,404$ | $1,291,890,737$ | $395,677,038$ |
| density | $2.52 \times 10^{-7}$ | $2.95 \times 10^{-7}$ | $4.85 \times 10^{-7}$ |
| reciprocity | 0.31 | 0.33 | 1.0 |
| max in-degree | $2,289,874$ | $1,822,999$ | N/A |
| max out-degree | 84,789,166 | $9,9813$ | N/A |
| max degree | N/A | N/A | 4,294 |

[1] In this paper we use the terms “user” and “node” interchangeable

[2] G+ assigns each user a 21-digit integer ID, where the highest order digit is always 1 (e.g., 100000000006155622736)

## 3 Overview of the Reciprocal Network

In this section, we first describe our methodology to extract the reciprocal network of G+. We then provide a brief overview of some global structural properties of the reciprocal network. Firstly, to derive the reciprocal network of G+, we proceed as follows: from $\Omega$, we extract the subgraph composed of nodes with at least one reciprocal edge. We label this new subgraph as $G$. However, $G$ is not a connected subgraph. Hence, we use BFS (breadth-first-search) to extract its *largest connected component* (LCC); we label this new subgraph as $H$. In this paper, we consider this subgraph $H$ as the "reciprocal network" of G+[3]. It consists of 40,403,216 nodes and 395,677,038 edges, with a density of $4.85 \times 10^{-7}$, slightly larger than the density of $\Omega$. The main statistics of $H$ are listed in the last column in Table 1.

Figure 2 shows the complementary cumulative distribution function (CCDF) of the degrees of nodes in $H$ – we note that they represent the *mutual* degrees or reciprocal degrees of the same nodes in $\Omega$. For comparison, we also plot the CCDFs of the in-degrees and out-degrees for these nodes in $\Omega$. We can see that these curves have approximately the shape of a power law distribution. The CCDF of a power law distribution is given by $Cx^{-\alpha}$ and $x, \alpha, C > 0$. By using the tool in [18, 19], we estimate the exponent $\alpha$ that best models each of our distributions. We obtain $\alpha = 2.72$ for mutual degree, $\alpha = 2.41$ for out-degree and $\alpha = 2.03$ for in-degree distributions. We observe that the mutual degree and out-degree distributions have similar x-axis range and the out-degree curve drops sharply around 5000. We conjecture that this is because G+ maintains a policy that allows only some special users to add more than 5000 friends to their circles [8]. The observed power-law trend in the distributions implies that a small fraction of users have a disproportionately large number of connections, while most users have a small number of connections – *this is characteristics of many social networks.*

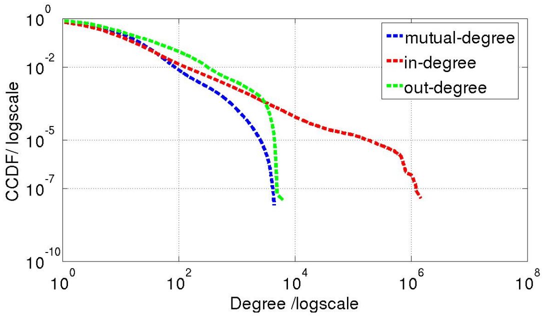


**Fig. 2.** Degree distributions for subgraph $H$

[3] It contains more than 90% of the nodes with at least one reciprocal edge in G+. Hence, our analysis of the dataset is eventually approximate.

## 4 Extracting the Core Graph of the Reciprocal Network

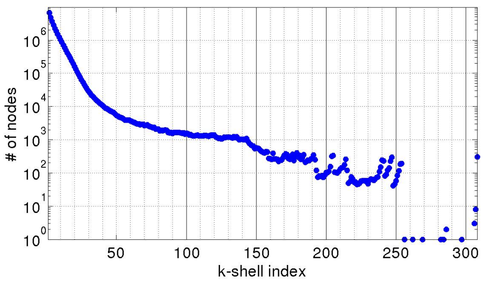
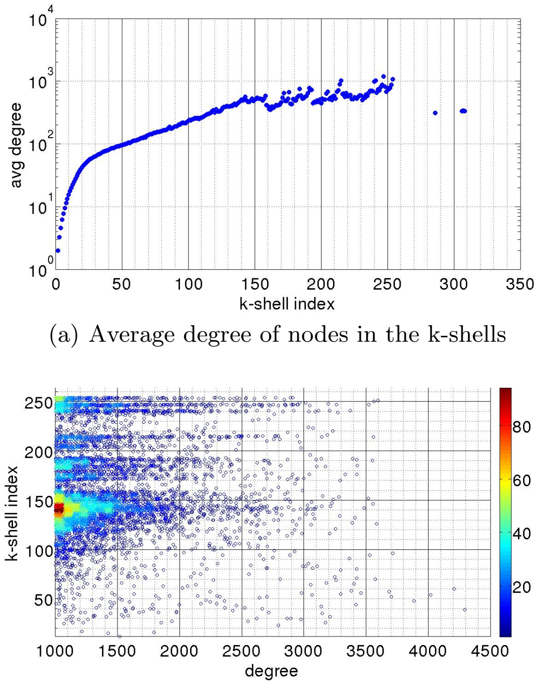


**Fig. 3.** The number of nodes belonging to the $k$-shell as $k$ varies from 1 to 308

Given its massive size (with more than 40 million nodes and nearly 400M edges), we apply a modified version of the *k-shell decomposition* method [13] to prune nodes and edges of *sparse* subgraphs that are likely to lie at the "periphery" of the G+ reciprocal network. K-shell decomposition is a classical graph decomposition technique which has been used as an analysis and visualization tool to extract and study the "core" structure of complex networks, such as that of the Internet AS graph [13]. The classical k-shell decomposition method works as follow: a) first, remove all nodes in the network with degree 1 (and their respective edges) – these nodes are assigned to the 1-shell; b) more generally, at step $k = 2, \ldots$, remove all nodes in the remaining network with degree $k$ or less (and their respective edges) – these nodes are assigned to the $k$-shell; and c) the process stops when all nodes are removed at the last step – the highest shell index is labelled $k_{max}$. At the end of the k-shell decomposition process, each node $v$ is assigned with a unique *k-shell index*, denoted by $shell(v)$ (whereas we use $deg(v)$ to denote the degree of $v$ in the network). The network can be viewed as the union of all $k_{max}$ shells, and for each $k$, we define the $k$-core as the union of all shells with indices larger or equal to $k$.

Clearly, for a node to belong to the $k$-core (thus $shell(v) \geq k$), it must have at least degree $k$, i.e., $deg(v) \geq k$. However, $deg(v) \geq k$ is not sufficient to guarantee it to belong to the $k$-core. For example, a node $v$ with only neighors of degree 1 (i.e., $v$ is the root of a star structure) belongs to the 2-shell, i.e., $shell(v) = 2$, no matter how high its degree is. On the other hand, it is easy to see that if a node $v$ is part of a clique of $k$ nodes, then $shell(v) \geq k$. However, a node $v$ does not need to be part of a $k$-clique to have $shell(v) \geq k$. Consider a tree $T$ of $n$ nodes (the sparsest graph with $n$ nodes). We can in fact provide a complete characterization of nodes in $T$ to have $shell(v) \geq k$ in a recursive manner: for $v$ to have $shell(v) \geq k$, it must have at least $k$-neighbors $u$'s with $shell(u) \geq k - 1$ – this characterization also applies to a general graph. We see that in the case of a tree, nodes with higher $k$-shell indices must lie more at the "core" (i.e., the increasingly "denser" part) of the tree. For a general graph, however, a node with a high $k$-shell index may not lie at the "core" of the graph: it can be part of a large clique that is "isolated" on a periphery of a massive graph. In such a case, the large clique will break off from the "core" of the network (e.g., as represented by the largest connected component remaining in the $k$-core) in the early stage of the $k$-shell decomposition process.



(a) Average degree of nodes in the k-shells

(b) K-shell distribuition of the nodes with $deg(v) \geq 1000$

**Fig. 4.** Degree distributions for nodes in the k-shells

We apply the k-shell decomposition method to the G+ reciprocal network. We find that the $k_{max} = 308$, and the $k_{max}$-core is a clique of size 298 nodes (the maximum clique in the G+ reciprocal network). Figure 3 shows the number of nodes belonging to the $k$-shell as $k$ varies from 1 to 308: we see that 99% of the nodes in our network fall in the lower k-shells (from $k = 1$ to 100). This is not surprising, as the majority of the nodes in our network have degree less than 100. Figure 4(a) shows the average degree of nodes in the $k$-shell, whereas in Fig. 4(b) we zoom in on nodes with $deg(v) \geq 1000$, and illustrate how they distribute across various $k$-shells. We see that while a large portion of high-degree nodes belong to higher $k$-shells, in fact the highest degree nodes belong to lower

$k$-shells, suggesting that they do not lie at the "core" of the G+ reciprocal network.

Figure 5 shows the size of the largest as well as those of the 2nd, 3rd and 4th largest connected components in the $k$-core, as $k$ varies from 1 to 308. We note that at step $k = 121$, a small subgraph containing the maximum clique (of size 298) breaks off from the largest connected component which desolves after $k = 253$, whereas this subgraph containing the maximum clique persists after $k = 252$ and becomes the largest component, and at $k_{max} = 308$, we are left with the maximum clique plus 10 additional nodes that are connected to the maximum clique. Closer inspection of nodes in the maximum clique reveals that its users belong to a single institution in Taiwan, forming a close-knit community where each user follows everyone else. We see that directly applying the standard k-shell decomposition to the G+ reciprocal network produces a clique of size 298, which we believe is unlikely to be the "core" of the G+ reciprocal network.

In order to extract a meaningful "core" of the G+ reciprocal network, we therefore modify the standard k-shell decomposition method to stop the process earlier using the following criterion: we terminate the process at $k_C$ when the largest connected component breaks apart in two or more pieces where each contains a dense subgraph (e.g., a clique of size $q \gg k_C$, here we use a threshold of $q = 200$). Applying this criterion, we terminate the $k$-shell decomposition at $k_C = 120$, which yields the $k_C$-core graph with $k_C = 120$: this core graph $G_{120}$ has 51,189 nodes and 7,133,227 edges, with an average degree of 139.4 and a density of approximately 0.0054, which is much smaller than that of the reciprocal network $H$ as a whole. Figure 6 shows the degree distribution of the nodes in the 120-core graph (note that degree here refers to that of a node in $G_{120}$, the 120-core graph after the $k_C$th shell decomposition process, it is *not* the (original) degree of the node in the G+ reciprocal network). From Fig. 4(a) and Fig. 4(b), we see that $G_{120}$ is comprised of many nodes with (original) high degrees in the G+ reciprocal network, with an average (original) degree of roughly 500.

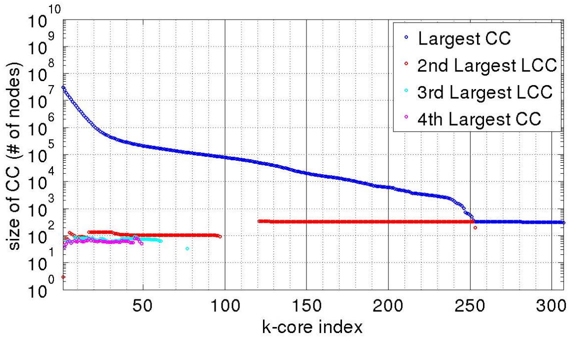


**Fig. 5.** The size of the largest as well as those of the 2nd, 3rd and 4th largest connected components in the k-core, as k varies from 1 to 308.

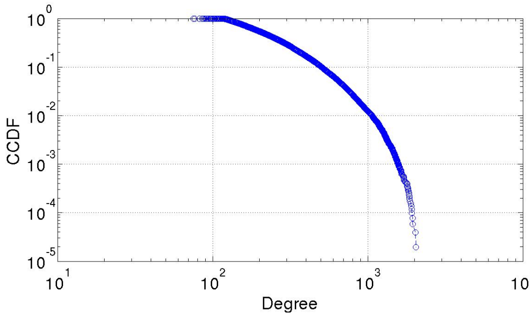


**Fig. 6.** Degree distribution for nodes in subgraph $G_{120}$

## 5 Clique Percolation Analysis & Core Clique Graph

In this section, given the dense core subgraph $G_{120}$, we want to extract the minimal set of the largest maximal cliques that cover every node in $G_{120}$ and use these cliques to build a new (hyber)graph that provides a higher-level representation of the core structure of the G+ reciprocal network. To achieve this, we proceed as following:

First, we implement an algorithm to extract a maximal clique containing a given node in a network. The algorithm is a variation of the popular Bron-Kerbosh algorithm [20]. Hence, we name it Simplified Bron-Kerbosh (SBK) and it is described in algorithm 1 – the parameter $t$ is used to set an upper bound on the size of the recursion tree. Then, we develop a procedure to extract the minimal set of the largest maximal cliques that cover every node in a given graph (algorithm 2). The resulting set of cliques returned from this procedure is always guaranteed to contain at least an unique node per clique. We apply this procedure to subgraph $G_{120}$. We obtain 37,005 maximal cliques with an average clique size of 22.26 nodes. Figure 7 shows the clique size distribution.

**Algorithm 1** Simplified Bron-Kerbosh (SBK)

```
$u$ : pivot vertex
$R$ : currently growing maximal clique
$P := N[u]$: set of neighbors of vertex u
$t$ : size of the largest clique allowed
SBK$(R, P, u, t)$
if $P := 0$ or $R := t$ then
    Report R as a maximal clique
else
    Let $u_{new}$ be the vertex with highest number of neighbors in P
    $R_{new} := R \cup \{u_{new}\}$
    $P_{new} := P \cap N[u_{new}]$
    SBK$(R_{new}, P_{new}, u_{new}, t)$
```

**Algorithm 2** Extract Minimal Set of Maximal Cliques from a Graph

```
1: procedure EMC($G(V, E)$)
2:     construct a set $W$ and $W := V$
3:     construct a ordered list $S$ of the nodes in $V$ based on their degree (decreasing order)
4:     select the first item in $S$, vertex $i$, as the pivot
5:     apply the SBK algorithm using $i$ as the pivot vertex
6:     add the reported maximal clique $c_i$ containing $i$ to the clique set $C_{total} = [c_n, c_m, ..]$
7:     remove the nodes in $c_i$ from $W : W_j = W_i - c_i$
8:     select the next item in $S$, vertex $j$, as the next pivot vertex such that $j \notin C_{total}$ and repeat
   steps(5), (6) and (7) until $W = \emptyset$
```

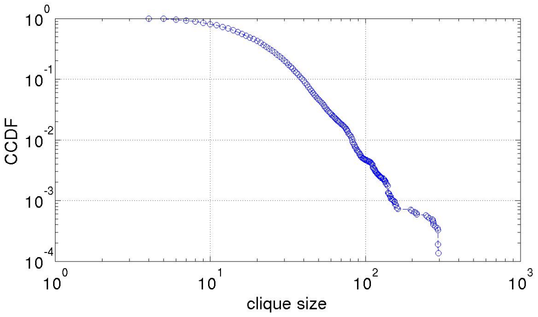
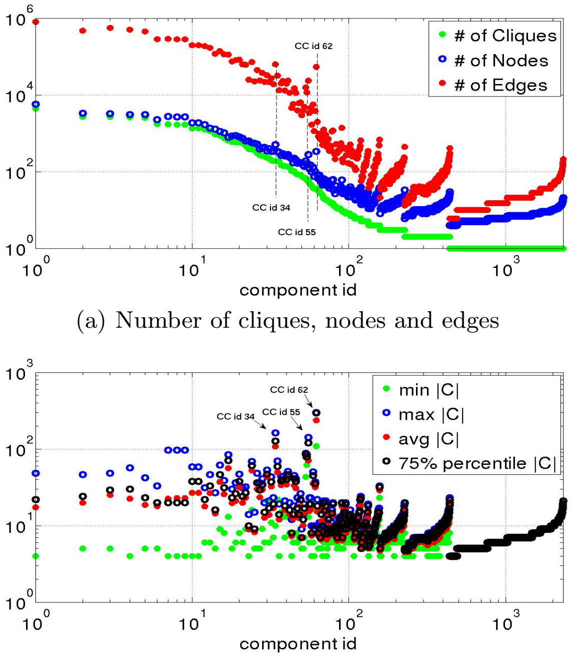


**Fig. 7.** Clique size distribution for subgraph $G_{120}$

Second, using the extracted 37,005 maximal cliques, we generate a new *directed* (hyber)graph, where the vertices are (unique) cliques of various sizes, and there exists a *directed* edge from clique $C_i$ to clique $C_j$ if more than half of the nodes in $C_i$ are contained in $C_j$, i.e., $C_i \rightarrow C_j$ if $(|C_i| \cap |C_j|)/|C_i| \geq \theta = 0.5$. We vary the parameter $\theta$ from 0.5 to 0.7, and find that it does not fundamentally alter the connectivity structure of the (hyper)graph of cliques thus generated. We remark that the maximal clique containing each node $v$ can be viewed as the most stable structure that node $v$ is part of. The directed (hyper)graph of cliques captures the relations among these stable structures each node is part of: intuitively, each directed edge in a sense reflects the attraction (or gravitational pull) that one clique (a constellation of nodes) has over the other. Hence this (hyber)graph of cliques provides us with a higher-level representation of the dense core graph of the G+ reciprocal network – how the most stable structures are related to each other. This procedure can be viewed as a form of clique percolation [14].

We find that this (hyper)graph of cliques comprises of 2,328 connected components (CCs). The largest component has 4,411 cliques, 5,697 nodes and 799,076 edges, while the smallest has 1 clique, 21 nodes and 210 edges respectively. We regard these connected components (CCs) as forming the *core communities* of the core graph of the G+ reciprocal graph: each CC is composed of either one single clique (such a CC shares few than half of its members with other cliques or CCs), or two or more cliques (stable structures) (where one clique shares at least half of its member with another clique in the same CC, thus forming a closely knit community). Figure 8(a) shows the distributions of these components in terms of the number of cliques, the number of nodes and the number of edges. We observe that for CC id's from 1 to 100 (which contains 30 or more cliques), there is a strong correlation between the number of cliques, nodes and edges: in general the connected components with the highest number of cliques also have the highest number of nodes and edges.



(a) Number of cliques, nodes and edges

(b) Clique size: maximum, minimum, average and 75% percentile

**Fig. 8.** Distributions of the connected components in the (hyper)graph of cliques

Figure 8(b) shows the maximum, minimum, average and 75% percentile of clique size for each *CC*. We observe that there is not a relationship between the number of cliques and their respective sizes in the *CCs*. We observe that most cliques have sizes between 10 and 100 nodes. There are largest *CCs* composed with a huge number of cliques of small size (e.g., *CC* ids from 1 to 10), whereas there are also small *CCs* composed with few number of cliques but with very large sizes (e.g. *CC* ids: 34, 55, and 62). We note also that there are a number of CCs which contain only one clique, but some of these cliques are of large size also.

## 6 Core Community Structure Analysis

In this section, we investigate the relationship between the connected components (CCs) in our (hyper)graph of cliques constructed in the previous section (Sect. 5), in particular the 65 largest CCs. First, regarding these CCs as the core community structures (a dense cluster of cliques) of the G+ reciprocal network, we define three metrics to study the relations among these CCs in the underlying G+ reciprocal network:

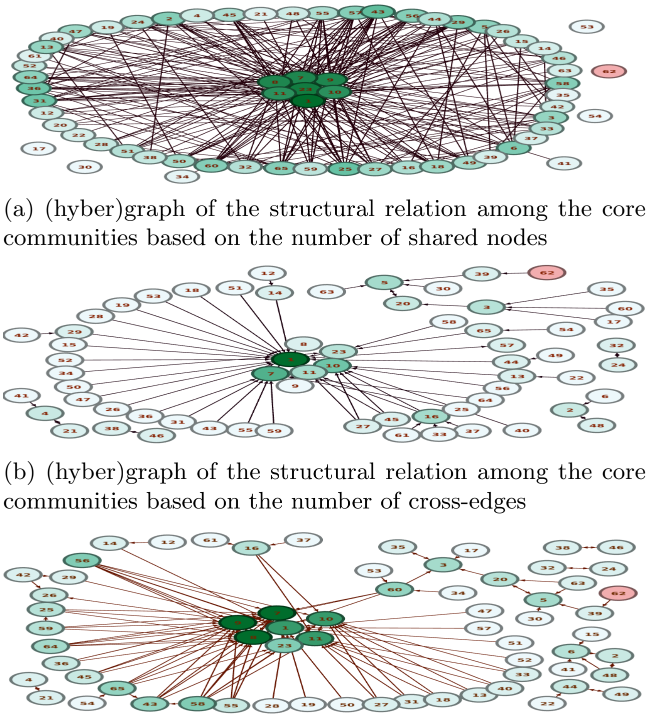
- *Shared nodes*: the number of nodes that $CC_i$ and $CC_j$ have in common
- *Shared neighbors*: the number of nodes in $CC_i$ that have at least one edge to a node in $CC_j$
- *Cross-edges*: the number of cross edges between $CC_i$ and $CC_j$

These metrics produce a set of new (hyber)graphs that succinctly summarize the (high-level) structural relations among the core community structures. They provide a big picture view of the core graph of the G+ reciprocal network and yield insights as to how it is formed. Figure 9(a) shows the (hyber)graph of the relationship between the components based on the number of shared nodes. We observe that there are seven CCs that lie at the center of this (hyber)graph through which the other CCs are most richly connected.

For the remaining two metrics, we observe that every $CC_i$ has at least one cross-edge and consequently one neighboring node with every other $CC_j$; thus the CC graph generated based on cross-edge or shared neighbors forms a complete graph – a clique. Hence, we focus our analysis on the strongest relationship between the $CCs$: for every $CC_i$, we extract the $CC_j$ that has the largest number of cross-edges with $CC_i$; likewise, for the neighboring nodes. Figure 9(b) shows the (hyber)graph of the relationship between the $CCs$ based on their number of "cross-edges": a node represents a $CC$ and a directed edge $CC_i \rightarrow CC_j$ implies that $CC_i$ has the largest number of cross edges to nodes in $CC_j$. Similarly, Fig. 9(c) shows the (hyber)graph of the relationship between the $CCs$ based on the number of "shared neighbors": a node represents a $CC$ and a directed edge $CC_i \rightarrow CC_j$ implies that $CC_i$ has the largest number of neighboring nodes with $CC_j$. These figures show that most $CCs$ have the largest number of cross edges and shared neighbors with the same seven $CCs$ identified in Fig. 9(a). Table 2 shows a summary of the statistics for the seven CCs, respectively. Based on these results, we conclude that there are seven subgraphs (core communities) comprising of dense clusters of cliques that lie at the center of the core graph of the G+ reciprocal network, through which other communities of cliques are richly connected. The 2,328 connected components (CCs) in the clique (hyper)graph form the core graph of the G+ reciprocal network, to which other nodes and edges that are part of sparse subgraphs on the peripherals of the network are attached.

We note in particular that in the periphery of our (hyber)graphs, we find a small CC composed with 35 of the largest cliques in the G+ reciprocal network. The average, minimum and maximum sizes of the cliques in this CC are 237, 109 and 298 – the latter is the maximum clique of the G+ reciprocal network. This

CC is highlighted by a “red circle” in the (hyper)graphs in Fig. 9. It shows this CC lies more at the outer ring of G+s dense core structure. As mentioned earlier in Sect. 4, the 298 users in this maximum clique of the G+ reciprocal network belong to a single institution in Taiwan where every user follows every other. The users in this clique also form close relations with many other users, forming 34 other cliques. Together, these 35 cliques form a close-knit community. However, we see that this community in fact does not lie at the very “center” – instead lies more at the outer ring – of the core graph of the G+ reciprocal network. Hence, we see that simply applying the conventional k-shell decomposition method to the G+ reciprocal network would yield the maximum clique in the G+ reciprocal network, but not its *core* structure. In constrast, the seven CCs mentioned above more likely lie at the “center” of the core graph of the G+ reciprocal network.



(a) (hyber)graph of the structural relation among the core communities based on the number of shared nodes

(b) (hyber)graph of the structural relation among the core communities based on the number of cross-edges

(c) (hyber)graph of the structural relation among the core communities based on the number of neighboring nodes

**Fig. 9.** (Hyper)Graphs for the core communities of the reciprocal network of G+

**Table 2.** Summary of the statistics for the seven components that lie at the center in the core graph of the reciprocal network of G+

| ID | # nodes | # edges | # c | min $\|c\|$ | max $\|c\|$ | avg $\|c\|$ | 75% percentile |
|---|---|---|---|---|---|---|---|
| 1 | 5,697 | 799,076 | 4,411 | 4 | 48 | 17.3 | 22 |
| 7 | 2,736 | 287,529 | 1,715 | 4 | 97 | 22.5 | 20 |
| 8 | 2,669 | 279,607 | 1,663 | 4 | 97 | 22.8 | 20 |
| 9 | 2,668 | 279,486 | 1,662 | 4 | 97 | 22.9 | 20 |
| 10 | 1,895 | 196,341 | 1,345 | 4 | 58 | 26.8 | 38 |
| 11 | 1,894 | 196,281 | 1,344 | 4 | 58 | 26.8 | 38 |
| 23 | 621 | 24,794 | 386 | 4 | 14 | 8.9 | 10 |

## 7 Conclusion

In this paper we have developed an effective three-step procedure to *hierarchically* extract and unfold the *core* structure of the reciprocal network of Google+. We first applied a modified version of the k-shell decomposition method to prune nodes and edges of sparse subgraphs that are likely to lie at the peripherals of the G+ reciprocal network. We then performed a form of clique percolation to generate a new *directed*) (hyper)graphs where vertices are maximal cliques containing the nodes in the dense "core" graph generated in the previous step, and there exists a directed edge from clique $C_i$ to clique $C_j$ if half of the nodes in $C_i$ are contained in $C_j$. We found that this (hyper)graph of cliques comprises of 2000+ connected components (CCs), which represent the the core "communities" of the G+ reciprocal network. Finally, we introduced three metrics to study the relations among these CCs in the underlying G+ reciprocal network: the number of nodes shared by two CCs, the number of nodes that are neighbors in the two CCs, and the number of edges connecting these neighboring nodes. These metrics produce a set of new (hyber)graphs that succinctly summarize the (high-level) structural relations among the core "community" structures and provide a "big picture" view of the core structure of the G+ reciprocal network and how it is formed. In particular, we found that there are seven CCs that lie at the center of this core structure through which the other CCs are most richly connected. As part of ongoing and future work, we will develop a more rigorous characterization of the core graph of the G+ reciprocal network based on the (modified) k-shell decomposition, and provide a more in-depth analysis of the (hyber)graph structures of the clique core graph and the (high-level) structural relations among the core "community" structures. We also plan to apply our method to other directed OSNs such as Twitter.

**Acknowledgments.** This research was supported in part by DoD ARO MURI Award W911NF-12-1-0385, DTRA grant HDTRA1- 14-1-0040 and NSF grant CNS-1411636. We thank the authors of [6] for the datasets.

# References

1. Gong, N.Z., Xu, W.: Reciprocal versus parasocial relationships in online social networks. Soc. Netw. Anal. Min. 4(1), 184197 (2014)
2. Garlaschelli, D., Loffredo, M.I.: Patterns of link reciprocity in directed networks. Phys. Rev. Lett. 93, 268–701 (2004)
3. Jiang, B., Zhang, Z.-L., Towsley, D.: Reciprocity in social networks with capacity constraints. In: KDD 2015, pp. 457–466. ACM (2015)
4. Hai, P. H., Shin, H.: Effective clustering of dense and concentrated online communities. In: Asia-Pacific Web Conference (APWEB) 2010, pp. 133–139. IEEE (2010)
5. Gong, N.Z., Xu, W., Huang, L., Mittal, P., Stefanov, E., Sekar, Song, D.: Evolution of the social-attribute networks: measurements, modeling, and implications using Google+. In: IMC 2015, pp. 131–144. ACM (2015)
6. Gonzalez, R., Cuevas, R., Motamedi, R., Rejaie, R., Cuevas, A.: Google+ or Google-? dissecting the evolution of the new OSN in its first year. In: WWW 2013, pp. 483–494. ACM (2013)
7. Kwak, H., Lee, C., Park, H., Moon, S.: What is twitter, a social network or a news media? In: WWW 2010, pp. 591–600. ACM (2010)
8. Magno, G., Comarela, G., Saez-Trumper, D., Cha, M., Almeida, V.: New kid on the block: exploring the Google+ social graph. In: IMC 2012, pp. 159–170. ACM (2012)
9. Mislove, A., Marcon, M., Gummadi, K.P., Druschel, P., Bhattacharjee, B.: Measurement and analysis of online social networks. In: IMC 2007, pp. 29–42. ACM (2007)
10. Wolfe, A.: Social network analysis: methods and applications. Am. Ethnologist 24(1), 219220 (1997)
11. Jamali, M., Haffari, G., Ester, M.: Modeling the temporal dynamics of social rating networks using didirectional effects of social relations and rating patterns. In: WWW 2011, pp. 527-536. ACM (2011)
12. Li, Y., Zhang, Z.-L., Bao, J.: Mutual or unrequited love: identifying stable clusters in social networks with uni- and bi-directional links. In: Bonato, A., Janssen, J. (eds.) WAW 2012. LNCS, vol. 7323, pp. 113–125. Springer, Heidelberg (2012)
13. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y. and Shir, E.: A model of Internet topology using k-shell decomposition. PNAS 104, 11150-11154 (2007).
14. Palla, G., Dernyi, I., Farkas, I. and Vicsek, T.: Uncovering the overlapping community structure of complex networks in nature and society. Nature, 435(7043), 814-818 (2005).
15. Schiberg, D., Schneider, F., Schiberg, H., Schmid, S., Uhlig, S., Feldmann, A.: Tracing the birth of an OSN: social graph and profile analysis in Google+. In: WebSci 2012, pp. 265–274. ACM (2012)
16. Google+ Platform, http://www.google.com/intl/en/+/learnmore/
17. Google+, http://en.wikipedia.org/wiki/Google+
18. Clauset, A., Shalizi, C. R., and Newman, M. E. J.: Power-law distributions in empirical data. SIAM Rev. 51, 661–703 (2009)
19. Fitting Power Law distribution, http://tuvalu.santafe.edu/ aaronc/powerlaws/
20. Cazals, F. and Karande, C.: A note on the problem of reporting maximal cliques. Theoretical Computer Science, 407(1), 564-568 (2008).