# From Caesar to Twitter
## Structural Properties of Elites and Rich-Clubs

Chen Avin[1], Zvi Lotker[1], Yvonne-Anne Pignolet[2], Itzik Turkel[1]
[1] Ben Gurion University of the Negev, Be'er-Sheva, Israel
[2] ABB Corporate Research, Baden, Switzerland
{avin, zvilo, turkel}@cse.bgu.ac.il, yvonne-anne.pignolet@ch.abb.com

***Abstract*—In many societies there is an *elite*, a relatively small group of powerful individuals that is well connected and highly influential. Since the ancient days of Julius Caesar's senate to the recent days of celebrities on Twitter, the size of the elite is a result of conflicting social forces competing to increase or decrease it. In this paper we formulate these forces as axioms and study their equilibrium and other properties of elites in social networks and complex systems.**

**Our findings indicate that elite properties such as a *size* of $\Theta(\sqrt{m})$ (where $m$ is the number of edges in the network), disproportionate *influence*, *stability* and *density* are universal and should join an increasing list of common phenomenon that complex systems share such: "small world", power law degree distributions, high clustering, etc. As an approximation for the elite we study the subgraph formed by the highest degree nodes, also known as the rich-club. We analyze the structural properties of the $k$-rich-club of nine existing complex networks and three theoretical models systematically, where the $k$-rich-club is the subgraph induced by the $k$ nodes with the highest degree in the network. In all real-life networks we observe similar elite properties for rich-clubs consisting of around $\sqrt{m}$ nodes, however, none of the theoretical models we analyzed captures all the elite properties, and thus they should be either adjusted or extended to address these findings.**

## I. Introduction

In the past few decades, the study of the structure of complex systems and social networks revealed some of the universal properties they share. Some of the basic properties that these networks exhibit are: short average length (i.e., "six degrees of separation"), a high clustering coefficient, a heavy-tailed degree distribution (e.g., scale-freeness), navigability and more recently densification and a shrinking diameter [34], [1], [21], [27]. These empirical findings led to a variety of random graph models that are trying to capture properties and the evolution of social networks. In turn, the models are used to predict and better understand the basic mechanisms that govern these networks. Some of the most popular models are the Preferential Attachment model (BA) [1], Kleinberg's small-world model [12] the Copy model [16], the Forest Fire model [21] and more recently the Affiliation Networks model [18].

When "new" universal properties are found by empirical measurements, the existing models need to be rechecked with regard to them, and if necessary to be improved or replaced by new ones. In the current work, we draw attention to another basic and important phenomenon of (social) networks structure: the existence of an *elite*. In the Cambridge Dictionary the elite is defined as:

> "*The richest, most powerful, best educated or best trained group in a society.*"

Other definitions (e.g., Wikipedia) emphasize in addition that the elite group is "small" and "well-connected".

Intuitively, the best candidates for the elite group are nodes with the highest degree in the network, also called the *rich-club* [37]. These nodes are well-known to exist following the scale-free nature of complex systems and the power-law degree distribution that enforce such "superstars"; nodes with degree well above the average degree (aka "hubs" and "connectors") [6]. Previous research on the rich-club phenomenon already demonstrated the existence of some interesting properties like the tendency of high degree nodes to be well connected among each other [4], [24], [37]. The importance of the rich-club with respect to the whole network was considered in [36] which shows that the rich-club connectivity has a strong influence of the assortativity and transitivity of a network (i.e., whether connections between nodes of similar degree are more likely and how many triangles occur). Based on these findings, the rich-club can be seen as the elite of a complex network due its influence on the rest of the network. We refer to Section V for a discussion of related notions.

In this paper we perform a more systematic study on the structural properties of the *rich-club subgraph*; the graph that is induced by the highest degree nodes. This is motivated by several reasons. First of all, structural properties of the rich-club in a network may help understand the mechanisms and behavior of the whole network. Second, knowledge on these properties might facilitate the construction of new algorithms and heuristics to solve important problems, that are hard to solve on general graphs. And finally, another advantage of studying the rich club is its small size compared to the size of the complete network. Social networks can comprise billions of users and even more edges which makes an analysis more difficult. A smaller, yet important set of nodes can be analyzed with more sophisticated tools such as eigenvalue decompositions and flow computations, which might not be feasible for the whole network.

The novel question we ask about the rich-club (and more generally the elite), is maybe the most basic one: "What is the size of the elite in complex networks?" Moreover, is the elite size a universal property of complex networks, similar to other universal properties we mentioned earlier? Our results indicate that this is the case, and the elite size, as well as others of its structural properties, are universal.

We observe that the size of the elite is determined by conflicting forces. Forces that try to increase the elite size versus forces that try to reduce its size. As a motivation for this tension and process we were inspired by some known anecdotes from human history. Ancient Roman history provides an example for forces that aspire to increase the size of the Roman elite, namely, the senate [10], [15]. When established around 750 BC, the senate included representatives of the first 100 families. When the population grew, it was extended, reaching 300 members in 509 BC, and then 600 senators in 80 BC. It was Julius Caesar who finally increased the senate to around a thousand senators. Interestingly, this number is about the square-root of the Rome population at the time, one million. On the other hand, the French Revolution 1789–1799, provides an example for a society that reduced its elite size. During the Reign of Terror from 1793 to 1794 thousands of French elite members, the nobility, were executed by the guillotine [8].

Going back to our days, we make an effort to formalize some of these social forces in terms of graphs and networks and to derive the elite size at equilibrium. A complex network is modeled as a graph $G = (V, E)$ with $n = |V|$ nodes connected by a set of (directed) edges $E$, $m = |E|$. These edges represent a relation between two nodes, such as friendship, citations, following on Twitter, etc. We define the $k$-rich-club of a network to be the subgraph induced by the $k$ highest-degree nodes. As it turns out, this simple definition leads to some interesting observations when investigating the structure of the inter-connectivity among the highest degree nodes and their interactions with the rest of the network for a growing number of elite members, i.e., for $k$ starting at 1 to the total number of nodes. Our results show that the $k$-rich-club has different structural properties than the whole network it belongs to and suggest a set of measurements to quantify the "power" of the $k$-rich-club. To the best of our knowledge, the set of properties we study (see next) have not been analyzed for growing rich-clubs before and we believe that future models for social networks should capture the universal properties of the rich-club because of its role and significance.

### *Summary of our Results*

**Elite size**: We take an axiomatic approach to conclude that the size of the elite is in the order of $\sqrt{m}$, where $m$ is the number of edges in the network. This result follows from assuming that the elite is *influential*, *stable* and either of *minimum-size* or *dense* (see Section III).

We then measured a variety of parameters for rich-clubs of growing size, providing empirical evidence that for a $\sqrt{m}$-rich-club the following statements hold for its structure and its interaction with the whole network[1].

**Inner Structure**: (i) The induced subgraph of the $\sqrt{m}$-rich-club of existing social networks is *dense*, in particular much denser than the whole network. (ii) The largest connected component of this subgraph contains almost all rich-club nodes. (iii) The average degree of the $\sqrt{m}$-rich-club in its induced subgraph is significantly higher than the average degree of the whole networks. Note that these findings are NOT a mere consequence of the fact that the rich-club contains the highest degree nodes (cf. to networks with the same number of edges generated according to some complex network models discussed later).

**Influence**: The elite has a "disproportionate" power toward the society. In graph terms, a significant constant fraction of nodes outside the $\sqrt{m}$-rich-club have a neighbor in the $\sqrt{m}$-rich-club. Related to this is the fact that the size of the cut between the $\sqrt{m}$-rich-club and the rest of the network is a significant constant fraction of all edges in the network.

**Stability**: The elite is stable against "outside" pressure from the society. In graph terms, the ratio of the outgoing edges from elite to the inner edges of the elite is constant.

**Symmetry**: In directed networks the $\sqrt{m}$-rich-club is significantly more symmetric than the whole network.

**Evolution**: There is a high correlation between the high degree nodes and the *seniority* of members in the networks. Note that while some models predict this well, others do not.

Some of the above properties might have been known on a anecdotal level or may seem obvious, however, they have not been measured together for growing rich-clubs and they cannot be explained by *only* considering the fact that the $k$-rich-club contains the highest degree nodes. It does not hold for arbitrary networks that the structure of the $k$-rich-club has these properties. In order to demonstrate this, we compare our findings on real-world data to the properties the popular Erdös-Renyi model, the Barabási-Albert model and the Affiliation networks model exhibit. While there are some similarities, unfortunately these models fail to produce networks with a rich-club featuring *all* the properties found in real networks. Related and additional shortcomings have been pointed out for these and other models in previous work on the rich club phenomenon together with the need to devise improved models capturing this [35].

In the next section we introduce the notions, complex networks and models used in this paper, followed by an axiomatic derivation of the elite's size. After presenting our measurement results in more detail and reviewing related work, we discuss our findings and some major open questions raised by them in Section VI.

## II. Datasets and Models

Today several popular online social networking sites like Facebook, Twitter, Flickr, YouTube, Orkut, and LiveJournal exist. These networking sites are based on an explicit user graph to organize, locate, and share content as well as contacts. In many of these sites, links between users are public and can be crawled automatically. This allows researchers to capture and study a large fraction of the user graph. The obtained data sets present an ideal opportunity to measure and study online social networks at a large scale. Mislove et al. [26], [25],

[1] Part of our findings reinforce some of the claims of previous rich-club studies on additional real-world networks, other findings reveal new features. In this paper we study them together since they were not considered before in the context of a $\sqrt{m}$-elite size.

[31] have collected data from the most prominent online social networks and made them available to the research community. We used their data on Facebook, Livejournal, Orkut, Flickr, YouTube in addition to data provided by the Stanford Large Network Dataset Collection (http://snap.stanford.edu/data/) on Autonomous Systems (AS) and Wikipedia link graphs. Furthermore, we study the rich-club of Twitter [17] and a citation network (who cites whom) derived from DBLP and the ACM digital library.

To find out if the rich-club of real life complex networks is structurally different from arbitrary networks and to examine the rich-club of some well known graph models, we generate some graphs according to the the Erdös-Renyi random graph model, the Barabasi-Albert model and the Affiliation model. One of the first and most simple models for networks is the Erdös-Renyi (ER) random graph model [5]. In this model an edge between each pair of nodes exists with equal probability $p$, independently of the other edges. One model to generate scale-free graphs exhibiting some properties found in real networks is the *Barabási-Albert* (BA) model [1]. It captures growth and preferential attachment. More precisely it models the evolution of a social network, where nodes join the network and build links to existing nodes, based on their degree. The higher the degree of a node, the more likely it is to attract new nodes to connect to it. The network starts as an initial network of $m_0$ nodes. New nodes are added to the network one at a time. Each new node is connected to $m' \leq m_0$ existing nodes with a probability that is proportional to the number of neighbors that the existing nodes already have. Formally, the probability $p_i$ that the new node is connected to node $i$ is [1] $p_i = deg(i)/\sum_j deg(j)$, where $deg(i)$ is the degree of node $i$. In this work we adopt the convention $m_0 = m'$ and start with an initial network forming a complete graph (clique). Another model, based on a bipartite *affiliation* graph from which a social network is derived, was presented in [18]. The affiliation graph models the fact that people ("actors") are typically connected to other people via "societies" (e.g., schools we visited, streets we live in, companies we work for, etc.). The social network is obtained by folding the bipartite graph, i.e., by generating an (undirected) edge in the social network for paths of length two in the affiliation graph. The affiliation graph evolves by letting new actors and societies copy another node's neighbors with some probability in addition to preferential attachment edges based on the degree. For each of these models we produced graphs with 1 million nodes. The parameters we used were $p = 0.00002$ for the ER model, $m' = 10$ for the BA model, and $c_q = c_u = 2$ (the number of edges added in 1 evolution step), $s = 2$ (the number of edges added by preferential attachment) and $\beta = 0.5$ (how often the left/right side of the bipartite graph grows). We decided to use these models as most other models known to us are based on variations and combinations of these models. All data sets (with degree rank as node identifiers) that we used in this paper are publicly available by emailing avin@cse.bgu.ac.il.

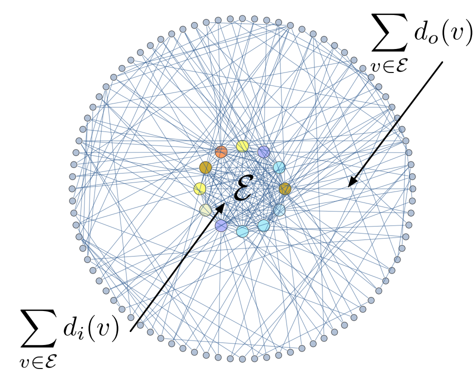


Fig. 1. Graphical demonstration of the elite and parameters in the Axioms. $m$ is the total number of edges in the graph, $\mathcal{E}$ is the elite set (colored nodes), $\sum_{v\in\mathcal{E}} d_o(v)$ is the total number of outgoing edges from the elite and $\sum_{v\in\mathcal{E}} d_i(v)$ is the total number of edges within the elite.

## III. Elite Size - An Axiomatic Approach

In this section we address one of the basic and most intriguing questions about the elite: what is its size? Many definitions of the elite indicate that the elite is a *small* group compared to the whole population. But what is the "right" size? What is small?

To answer these questions, and to explain our empirical results we take an axiomatic approach: we assume that an elite features some basic properties in order to maintain its power in the society and based on these we infer its size. We claim that the elite must be socially *stable* and *influential*. By adding either a *density* or *min-size* property we conclude that the elite size is in the order of $\Theta(\sqrt{m})$, where $m$ is the total number of edges. When the number of edges is proportional to the number of nodes $n$ then the elite size turns out to be $\Theta(\sqrt{n})$.[2]

More formally, let the set $\mathcal{E} \subset V$ denote the elite consisting of $|\mathcal{E}|$ nodes. For a node $v \in \mathcal{E}$ let $d_i(v)$ denote the internal degree of $v$ within $\mathcal{E}$, i.e., how many neighbors of $v$ belong to the set $\mathcal{E}$; analogously $d_o(v)$ denotes the number of neighbors of $v$ that are outside of the set $\mathcal{E}$, i.e., in $V \setminus \mathcal{E}$. Let $c_1, c_2, c_3$ be some constants. We postulate the following axioms for the elite set $\mathcal{E}$.

1. **Influence.** The number of out-going edges from the elite is a constant fraction of the total number of edges in the network.

$$\sum_{v\in\mathcal{E}} d_o(v) \geq c_1 \cdot m \tag{1}$$

for $0 < c_1 < 1$.

*Motivation*: This axiom captures the power and influence that are associated with the elites. In complex networks an edge can be interpreted as a source of influence, thus a powerful group must control a large fraction of edges in the network.

[2] While traditional research on social networks assumed that $m = \Theta(n)$, more recent models and observations[18], [21] show that $m = \omega(n)$, so $m$ and $n$ might differ in the order of magnitude.

| data | n | m | $\sqrt{m}$-rich-club | Influence - $c_1$ | Stability - $c_2$ | Density - $c_3$ |
|---|---|---|---|---|---|---|
| Youtube | 1138500 | 2989945 | 1729 | 35.60% | 7.90% | 5.70% |
| Facebook | 63732 | 817031 | 903 | 19.30% | 31.90% | 12.40% |
| Livejournal | 5204177 | 49163589 | 7011 | 9.50% | 20.30% | 3.90% |
| Orkut | 3072442 | 117174174 | 10824 | 10.20% | 26.20% | 5.40% |
| Flickr | 2302926 | 22830535 | 4778 | 34.80% | 39.60% | 27.50% |
| AuthorCitations | 85055 | 1234030 | 1110 | 31.80% | 24.40% | 15.60% |
| Wikipedia | 1870710 | 36473378 | 6039 | 41.80% | 6.00% | 5.00% |
| AS | 33560 | 75621 | 274 | 65.30% | 16.80% | 22.20% |
| Average | | | | 31.04% | 21.64% | 12.21% |
| STD | | | | 18.37% | 11.43% | 8.90% |
| ER Model | 1000000 | 9974503 | 3158 | 1.10% | 0.50% | 0.00% |
| BA Model | 1000000 | 9973255 | 3158 | 11.20% | 5.30% | 1.20% |
| Affiliation Model | 1000000 | 32092651 | 5665 | 11.60% | 220.50% | 51.30% |

TABLE I
BASIC PROPERTIES OF THE EXAMINED NETWORKS AND MODELS (# OF NODES, # OF EDGES) AND THE AXIOM CONSTANTS $c_1, c_2, c_3$ WHEN THE ELITE IS THE $\sqrt{m}$-RICH-CLUB. THE AVERAGE AND STANDARD DEVIATION ARE OF THE REAL NETWORKS ONLY. HIGHLIGHT CELL INDICATE A PROBLEM.

2. **Stability.** The number of edges within the elite is proportional to the number of out-going edges from the elite.

$$\sum_{v\in\mathcal{E}} d_i(v) \geq c_2 \cdot \sum_{v\in\mathcal{E}} d_o(v), \tag{2}$$

for $0 < c_2 < 1$.
*Motivation*: In order to adhere to its opinion, the elite must be able to resist "outside" pressure, otherwise individuals in the elite will change their option and will be influenced instead of being influential. Consider a node $v \in \mathcal{E}$ that makes decision based on a weighted majority of her friends. Since people in the elite are, by definition, more powerful (e.g., rich, educated, etc.) elite members' opinions count for more when $v$ consults it neighborhood. If we weigh friends within the elite with power 1, then the weight of the outside friends $w$ will be less than 1. $c_2$ represent the (average) power we associate with friends outside the elite in a stable case. Therefore we expect that $c_2 < 1$ in real networks.
3. **Minimum-Size/Compactness.** The number of elite members tends to be as small as possible.
*Motivation*: This axiom is based on some basic principals in science like *Principle of minimum energy*, *Principle of least action* and *Occam's razor*. Giving that other things are equal (such as influence and stability) the elite size will tend to be as small as possible. In social terms this can be motivated by the idea that if the elite holds a given revenue or power it will attempt to split it between as few members as possible.
4. **Density.** The elite is dense

$$\sum_{v\in\mathcal{E}} d_i(v) \geq c_3 \cdot \binom{|\mathcal{E}|}{2} \tag{3}$$

for $0 < c_3 < 1$.
*Motivation*: The goal of the density property is to capture the idea that the elite is a social "clique" where "everyone knows everyone". The density property holds when each member of the elite knows (on average) a constant fraction of the elite members.

Based on these axioms we can infer the size of the elite $|\mathcal{E}|$. First we show a lower bound.

*Claim 3.1:* If the elite satisfies Axioms 1 and 2 the size of the elite is: $|\mathcal{E}| = \Omega(\sqrt{m})$.

*Proof:* First note that $|\mathcal{E}|^2 > \sum_{v\in\mathcal{E}} d_i(v)$. Using Eq. (1) and (2) we get $|\mathcal{E}|^2 > \sum_{v\in X} d_i \geq c_1 \cdot c_2 \cdot m$ which implies $|\mathcal{E}| \geq \Omega(\sqrt{m})$. ■

It is important to note that Axioms 1 and 2 alone do not guarantee a *small* elite. Take for example a linear size elite, i.e., $|\mathcal{E}| = \Theta(m)$, of constant degree, e.g., a constant degree expander [9]. If additionally each member of the elite is connected to a constant number of nodes outside of the elite, the resulting elite is both stable and influential. In order to derive a small elite size we must assume additional axiom(s).We next show that assuming either Axiom 3 or 4, enable us to conclude an elite size in the order of $\sqrt{m}$.

*Theorem 3.2:* If the elite satisfies Axioms 1, 2 and 3 the size of the elite is: $|\mathcal{E}| = \Theta(\sqrt{m})$ and the elite is dense.

*Proof:* The upper bound of $|\mathcal{E}| = O(\sqrt{m})$ follows directly from Claim 3.1 and Axiom 3. Now assume the elite is not dense. Then $\sum_{v\in X} d_i = o(|\mathcal{E}|^2) = o(m)$. But this contradicts $\sum_{v\in X} d_i \geq c_1 \cdot c_2 \cdot m$, thus elite must be dense. ■

*Theorem 3.3:* If the elite satisfies Axioms 1, 2 and 4 then the size of the elite is: $|\mathcal{E}| = \Theta(\sqrt{m})$ and the elite is compact.

*Proof:* As before, $|\mathcal{E}|^2 > \sum_{v\in\mathcal{E}} d_i(v)$ and using Eq. (1), (2) and (3) we get $|\mathcal{E}|^2 > c' \cdot m \geq c'' \cdot |\mathcal{E}|^2$ for some constants $c', c''$. Hence it must hold that $|\mathcal{E}| = \Theta(\sqrt{m})$ which means the elite is compact (i.e., of minimum possible size when assuming Axiom 1 and 2 ). ■

An important note is in place: the Axioms and results above hold for **undirected** networks only. For **directed** networks like Twitter, a more general treatment is needed and we leave this for future work.

In context of the above results we would like to mention the pioneering work of Linial et. al. [23] and Peleg [29] where a notion related to elites is discussed, namely coalitions (subsets of nodes) that dominate neighborhood majority votings, so-called monopolies. Among the results regarding monopolies are lower bounds for the size of these coalitions, namely $\Omega(\sqrt{m})$ nodes are necessary to control the outcome of all neighborhood majority votings of general graphs and the number of edges within the monopoly has to be in the order of $m$. Interestingly, we use a different perspective but our Axioms 1 and 2 lead to the same result for a powerful elite.

In the next section we present measurements to support our claim that the elite size in social networks is of size $\Theta(\sqrt{m})$. We show that for existing networks a rich club of size $\Theta(\sqrt{m})$ is both stable in influential, where this is not the case for standard models like the BA and Affiliation model. As discussed earlier we will focus on the rich-club of different sizes as an approximation for the elite of the network.

## IV. Measurements on Real Data and Models

We studied a variety of parameters for rich-clubs of growing size for nine real networks and three theoretical models (Erdös-Renyi (ER) model, Barabási-Albert (BA) model and the Affiliation model). We put our focus on $k$-rich-clubs for $k$ in the order of $\sqrt{m}$ to point out our claim that rich-clubs around this size satisfy our axioms. First we examine the constants $c_1, c_2, c_3$ relevant to our axioms properties of influence, stability and density, and then other relevant properties.

### A. Axiom constants - $c_1, c_2, c_3$

Table I gives a summary on basic properties of the networks under scrutiny such as the number of nodes and edges ($n$ and $m$ respectively) and the influence, stability and density constants ($c_1, c_2, c_3$, respectively) for the $\sqrt{m}$-rich club of each network. Gray cells indicate a problem in the model, as we discuss next: i) Influence - the influence constant fluctuates between $9\% - 66\%$ in the real networks with high variance, $c_1$ of the BA and Affiliation models both fall in this range, but clearly (and intuitively) in the ER model, the $\sqrt{m}$-rich-club is not influential enough. ii) Stability - All the constants for the real networks are below one as expected. For the Affiliation model $c_2 > 2, (220\%)$, which contradicts our Axiom. iii) Density - Real networks show a density constant between $4\%$ and $28\%$ while all three models are problematic, ER and BA are too sparse and the Affiliation model is too dense. We discuss these findings in more detail later.

We now take a broader view of the results and check the above constants for an increasing size of $k$-rich-club from $k = 1$ to $k = n$. In order to compare networks of different sizes we use plots where the $x$ axis (the rich-club size, $k$) is normalized to $[0, 1]$. To focus on small $k$-rich-clubs, the $x$-axis describes the rich-club size for growing roots of the network size $n$, i.e., at $x \in [0, 1]$ the measurement point for the $n^x$-rich-club, in particular at $x = 0.5$ we have the values for the $\sqrt{n}$-rich-club. We emphasize the size of the $\sqrt{m}$-rich-club by adding a large dot at its location which differs for each network. In other figures (usually in a smaller size) we present the result on a linear $x$-axis where $k = nx$ which demonstrates that

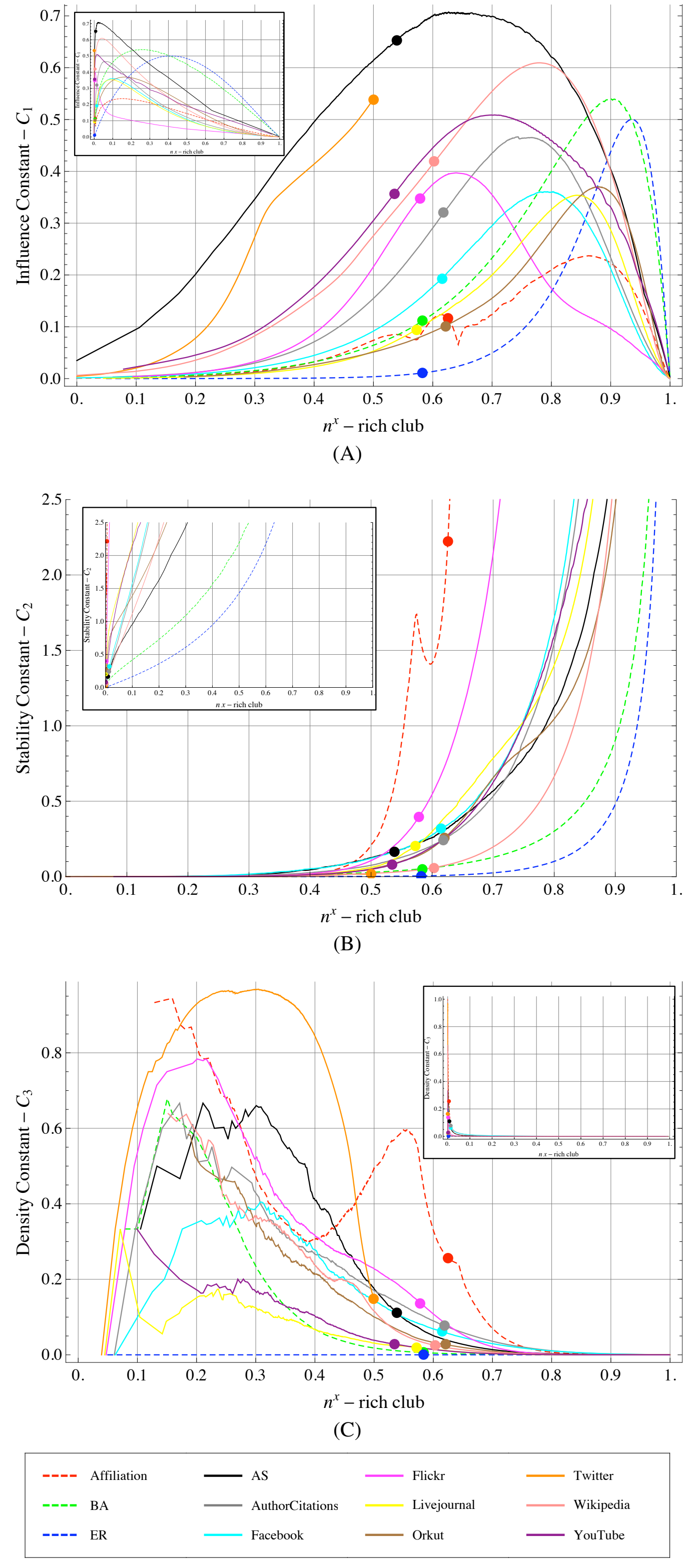


Fig. 2. Three graphs that show the values for influence, stability and density from the axioms on nine real networks (solid lines) and three models (dashed lines) for $n^x$-rich-clubs. (A) the influence constant - $c_1$, (B) the stability constant - $c_2$ and (C) the density constant - $c_3$. The dotes indicate where the $\sqrt{m}$-rich-club is located and at $x = 0.5$ the value for the $\sqrt{n}$-rich-club is depicted. The small figures are the same but with a linear scale, i.e., $k = nx$.

interesting phenomena occur for a very small $k$ compare to the network size (i.e., in the order of $\sqrt{m}$).

Fig. 2 (A), (B) and (C) presents the results for influence, stability and density of growing rich-clubs, respectively. Regarding influence - $c_1$, we observe some similar characteristics for all networks: i) influence increases monotonically until the maximum influence which is achieved at a rich-club size much larger than $\sqrt{m}$. ii) the constant $c_1$ is bound away from 0 at $\sqrt{m}$ (except in the ER model). Moreover, influence in most real networks is larger than in the models for low order $k$. Two extremes cases are the AS (Internet routers) network and Twitter. In particular Twitter is a directed network where the $k$-rich-club seems to have a much larger (directed) influence. Since we have data of Twitter only for a $k$-rich-club up to $k = \sqrt{n}$ not the whole scale is presented for this network.

For the stability constant $c_2$ we also observe similar behavior among the networks: stability monotonically increases with the rich-club size. Except for Twitter and the ER model the constant is clearly bounded away from 0 for $\sqrt{m}$-rich-club.

Recall that our axiom results about the constants hold (and are well-defined) for the undirected case only, whereas Twitter is a highly asymmetric directed network. As noted when discussing Table I $c_2$ of the Affiliation model exceeds 1 for at the $\sqrt{m}$-rich-club (and much earlier) which contradicts the second axiom.

An important observation is that to increase both influence and stability a larger $k$-rich-club is better. In contrast this in not the case with density. The density of the $k$-rich-club exhibits the opposite behavior (common to all networks, but the ER model): the maximum density is achieved at a rich-club size significantly smaller than $\sqrt{n}$ and from there on the density decreases monotonically (except for the Affiliation model). So for the $k$-rich-club to be denser is must be smaller, while to be influential and stable its size needs to be larger.

We conjecture that these conflicting forces determine the "right" size of the elite. The above empirical results (strengthened by the axioms) indicate that the balance between these forces, or the equilibrium, is achieved when the rich-club is in the order of $\sqrt{m}$. Moreover none of the three theoretical models seems to capture all of the properties (influence, stability and density) in the right scales as shown by the real networks.

To make the claims about the model more formal we can state the following:

*Proposition 4.1 (ER model rich-club density ):* The expected number of edges in the $k$-rich-club of a ER graph is $o(k^2)$ for $p = o(1)$ and large enough $k$.

It is also not hard to show that the BA model does not fulfill the density requirement

*Theorem 4.2 (Barabasi-Albert model rich-club density):* The expected number of edges in the $k$-rich-club of a Barabasi-Albert graph is linear, i.e., $O(k)$.

*Proof:* No matter which nodes belong to the $k$-rich-club, each node has $m$ outgoing edges in the BA model. Hence the total number of edges within the rich-club cannot exceed $2k \cdot m$. As a consequence the rich-club is not dense if $k$ is not in the same order of magnitude as $m$. ■

A more challenging task from a theoretical point is to prove that the Affiliation model does not capture the properties of the rich-club of complex systems or social networks. We leave this as a conjecture in this work and provide more empirical evidence in the next section.

### B. Seniority

Besides a high degree, what other properties do the rich-club members have? It is known that there is a strong correlation between high degree and time of arrival to the network [1], [14], [17]. We call nodes that arrive early *senior* members of the network. We would like to point out the arrival order of rich-club nodes in the Affiliation model compared to Wikipedia. Fig. 3 shows that in the Wikipedia graph the members of the 10'000-rich-club are indeed mostly seniors, i.e., they arrived early (low $y$-axis value). On the other hand, Fig.3 exposes what we think can be a major problem in the Affiliation model. The figure shows that significant number of the 10'000-rich-club are non-senior members, i.e., there are many nodes that arrived late (high $y$-axis value) but have a very high degree (low $x$-axis value). This can be intuitively understood from the model: in the Affiliation model, a late comer (i.e., non-senior) node usually joins a popular affiliation in the copying process. Once it joined an affiliation its degree is (immediately!) at least the size of the affiliation. This leads to a situation where all members of the largest affiliation (of which many members are not senior) are part of the rich-club. We can clearly see these phenomena in Fig. 3. The nodes in the same "black wave" in the plot belong to the same affiliation.

### C. Maximum Sociability

Another measure for the structure and connectivity of the $k$-rich-club is the *sociability*. The sociability of a graph is its normalized average degree, i.e., for $k$-rich-club with $m_k$ edges among the rich-club members this value is $\frac{m_k/k}{\max_{1 \le k' \le n} m'_k/k'}$. For a graph of growing size the maximum sociability captures the size of the network at which its members are, on average, most socially involved (or influenced) in the community. As mentioned earlier the average degree of the BA model is the same for any $k$-rich-club and therefore its sociability level is more or less constant after 5% of the network size. In contrast, real social networks are significantly different with the maximum sociability achieved at a $k$-rich-club of size around $n^{0.6}$. This can be seen in Fig. 4, where the figure shows that the maximum is achieved at a small scale $k$-rich-club. Interestingly, all real social networks have a single peak for the maximum, this may indicate that this point is a good candidate to define the "right" size of the rich-club. An exception to this rule of thumb is the Wikipedia graph with two maxima. When examining Figures 4, we notice that the maxima occur before or after $k = \sqrt{m}$ in some networks. One possible explanation is that our data sets are not complete, i.e. some nodes and edges are missing, another one is that these networks are not in a balanced state, i.e., the elite will grow or shrink until an equilibrium is reached.

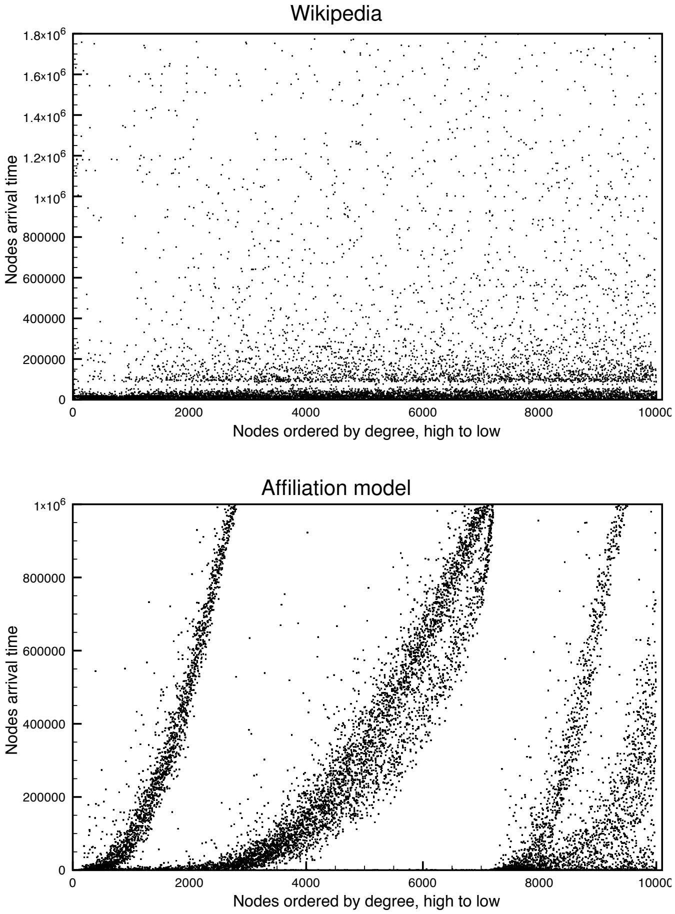


Fig. 3. (Top) Seniority in Wikipedia: high correlation between order of arrival ($y$-axis) and order of degree ($x$-axis). Most nodes depicted in this plot have arrived early and they belong to the 10'000-rich-club. (Bottom) Seniority in Affiliation model: many late comers (non-senior members) are part of the 10'000-rich-club

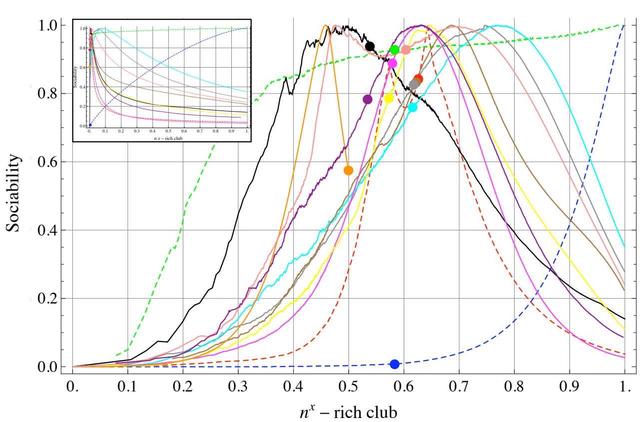


Fig. 4. Maximum Sociability: This graph depicts the number of rich-club edges divided by the number of rich-club nodes with the maximum normalized to one, for a $k$-rich-club with $m_k$ edges this value is $\frac{m_k/k}{\max_{1\leq k'\leq n} m'_k/k'}$. This ratio is equal to the average degree of the rich-club nodes.

### D. Elite Connectivity

In social networks, the largest connected component (LCC) typically covers almost all nodes of the network. However this does not imply that for any graph with a large LCC it must hold that the LCC of the $k$-rich-club contains almost all $k$ nodes. E.g., we found when analyzing the size of the LCC of the $\sqrt{m}$-rich-club reveals that almost all nodes in the rich-club of the social networks belong to the LCC. The same holds for the BA and Affiliation model. In the ER graph however, most rich-club nodes do not have any edges to other rich-club nodes, hence the rich-club is split into many separate components, most of them consisting of one node only.

| data | $\sqrt{m}$-rich-club | # comp | LCC |
|---|---|---|---|
| YouTube | 1729 | 9 | 1721 |
| Facebook | 903 | 1 | 903 |
| LiveJournal | 7011 | 16 | 6978 |
| Orkut | 10824 | 13 | 10812 |
| Flickr | 4778 | 1 | 4778 |
| Author Citations | 1110 | 1 | 1110 |
| Wikipedia | 6039 | 2 | 6038 |
| AS | 274 | 4 | 271 |
| ER | 3158 | 2888 | 3 |
| BA | 3158 | 1 | 3158 |
| Affiliation | 5665 | 1 | 5665 |

TABLE II
CONNECTIVITY TABLE OF $\sqrt{m}$-RICH-CLUB. THIS TABLE SUMMARIZES THE NUMBER OF CONNECTED COMPONENTS THE $\sqrt{m}$-RICH-CLUB, THE SIZE OF ITS LARGEST CONNECTED COMPONENT (LCC).

### E. Symmetry

In some networks the existence of an edge describes a reciprocal, symmetric relation between the two nodes involved (undirected network), whereas in other networks an edge from node $a$ to node $b$ (directed network) means that $a$ has a certain relationship with $b$ but not necessarily the other way around. Classically, sociologists make a distinction between directed networks and undirected networks when analyzing them. E.g., the first question of a decision tree for the analysis of cohesive subgroups on page 78 of [33] is “Is the network directed?”. The mathematical tools that are used differ depending on the answer, e.g., the notion of prestige (in-degree) does only apply to directed networks. On the other hand, in undirected graphs, degree centrality is used (see [32], Chapter 5). Clearly the directed graph model contains more information than its equivalent undirected version. However in many networks it is impossible or difficult to derive who initiated a relationship and/or what the direction of an edge is. For directed networks a natural question is whether the rich-club of the network is more symmetric than the rest of the network. Of our datasets the networks Wikipedia, Flickr, YouTube, Twitter and ER graph are directed. The average symmetric degree in the rich-club has a unique maximum in all three real networks. The maximum “ordinary” average degree of the rich-club is reached slightly after the maximum of the symmetric rich-club degree. Furthermore, it holds that the rich-club of the networks is more symmetric: the ratio between symmetric edges and all edges in the $k$-rich-club starts at almost 1 for $k = 2$ and then decreases rather quickly until reaching almost zero when $k$ approaches $n$. At around the maximum sociability ($k \approx \sqrt{n}$) the symmetric edges are still a significant fraction of all edges. In the ER graph model there are no symmetric edges which is not surprising for the chosen edge probability. Since the BA model and the Affiliation model are undirected they cannot help to explain or model the high symmetry within the rich-club.

In addition we counted the number of symmetric edges in the $\sqrt{n}$-rich-club of Twitter. In the following table we can see that 89% of the edges in the Twitter $\sqrt{n}$-rich-club are reciprocal, while in the whole Twitter network 22.1% of all edges are reciprocal [17].

| $m_{rc}$ | total | min | max | median | avg |
|---|---|---|---|---|---|
| directed | 5,537,573 | 0 | 3,778 | 656 | 852.07 |
| reciprocal | 4,952,210 | 0 | 3,238 | 512 | 762.00 |

When considering Twitter we notice that the $\sqrt{n}$-rich-club features especially high symmetry. One possible explanation for this is that the rich-club of Twitter is much larger than in the other networks and that this increases the social pressure on each of its members to increase the symmetry. Another explanation is that for Twitter many tools exist that help Twitter users to organize their tweets, followers and the users they are following. Among other features, some of these tools offer the functionality to add a new follower to the list of people their following. Presumably many of the high degree Twitter users apply such a software and "follow back" their followers. In order to find out if one of these theses is true, it is necessary to scrutinize data of other large networks and observe how the symmetry percentage changes with growing network size.

## V. Related Work

One of the first papers about the fact that the highest degree nodes are well connected examined the Autonomous Systems network [37] and coined the term rich-club coefficient for the ratio comparing the number of edges between nodes of degree greater than $k$ to the possible number of edges between these nodes. Colizza et al. [4] refined this notion to account for the fact that higher degree nodes have a higher probability to share an edge than lower degree vertices. They suggest to use baseline networks to avoid a false identification of a rich club. More precisely they propose to use the rich-club coefficient of random uncorrelated networks and/or the rich-club coefficient of network derived by random rewiring of edges while maintaining the degree distribution of the network. Weighted versions of the rich-club coefficient have been studied in [28], [30], [38] The question how the rich-club phenomenon manifests across hierarchies is studied in [24].

As identifying the most influential nodes in a network is crucial to understand its members behaviour, many other articles considered a variety of notions related to the elite and/or the rich-club. Mislove et al.[26] define the *core* of a network to be any (minimal) set of nodes that satisfies two properties: First, the core must be necessary for the connectivity of the network (i.e., removing the core breaks the remainder of the nodes into many small, disconnected clusters). Second, the core must be strongly connected with a relatively small diameter. As a consequence a core is a small group of well-connected group of nodes that is necessary to keep the remainder of the network connected. Mislove et al. use an approximation technique previously used in Web graph analysis, removing increasing numbers of the highest degree nodes and analyze the connectivity of the remaining graph. The core is thus the largest remaining strongly connected component. They observe that within these cores the path lengths increase with the size of the core when progressively including nodes ordered inversely by their degree. The graphs they study in [26] have a densely connected core comprising of between 1% and 10% of the highest degree nodes, such that removing this core completely disconnects the graph.

Another definition for a core can be found in [2]. Borgatti and Everett measure how close the adjaceny matrix of a graph is to the block matrix $\{\{1,1\},\{1,0\}\}$. This captures the intuitive conception that social networks have a dense, cohesive core and a sparse, unconnected periphery. Core/periphery networks revolve around a set of central nodes which are well-connected with each other, and also with the periphery. Peripheral nodes, in contrast, are connected to the core, but not to each other. On the other hand there are "clumpy" networks consist of two or more subgroups that are well-connected within each group but weakly connected across groups – like a collection of islands. When comparing networks with the same density, core/periphery networks have shorter average path lengths than clumpy networks. In addition to formalizing these intuitions, Borgatti and Everett devise algorithms for detecting core/periphery structures, along with statistical tests for testing a priori hypotheses[3].

The nestedness of a network represents the likelihood of a node to be connected to the neighbors of higher degree nodes. When examining this property, block modeling of adjacency matrices arranged by the degree of the nodes is also used. E.g., Lee et al [19] study such block diagrams for complex network models and they define a simple nestedness measure for unipartite and bipartite networks to capture the degree to which different groups in networks interact.

Apart from analyzing the most influential nodes, many articles have studied a wide range of properties of social networks. E.g., the networks YouTube, Flickr, Facebook, Wikipedia and LiveJournal have been analyzed in depth in [26], [25], [31].In addition there is a large body of papers studying information dissemination and path lengths [1], [11], [20], [7], and community structure [22], to name but a few examples.

## VI. Discussion and Open Problems

Reinforcing the claims of previous work on high degree nodes, our data analysis shows that many complex networks have a small subgraph which is much more dense than their complete network. In addition the structure of the whole network is influenced by this rich club. This can be exploited to find good candidate networks for the problem of finding the most dense subgraph (a NP-hard problem [13] on general graphs). One can apply the following procedure: Sort the nodes according to their degrees and choose the most dense subgraph among the subgraphs that containing the first $k$ highest degree nodes. We hope that this heuristic can be turned into an approximation algorithm once there are better models that capture rich-club properties of complex networks.

In addition we provide answers to the central question of how symmetry is spread among the edges of directed social networks. We show that edges inside the rich-club are much more symmetric than random edges that are not inside the rich-club. We can also see that in real complex networks most of the participants that are in the rich-club arrive early.

In order to make a step forward towards finding the "correct" size of the rich-club we use rich-club expansion to determine a subset which exhibits significant structural difference and influence to the rest of the network. This is closely related to the question of finding the elite of a complex network. Based on three axioms and on measurements performed on nine real-world networks, we conclude that the elite of a network consists of around $\sqrt{m}$ nodes and the $\sqrt{m}$-rich-club serves as a good approximation for the elite.

Unfortunately none of the existing models we examined are able to predict all the phenomena we describe. Hence, we support the quest raised in [35] for models capturing the main properties of complex networks and their elites continues, to be able to provide a better understanding of society and its communities.

## VII. Acknowledgements

We would like to thank the anonymous reviewers and arXiv readers for their suggestions for improvements and pointers to related work.

## References

[1] R. Albert and A. Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47–97, 2002.

[2] S. Borgatti and M. Everett. Models of core/periphery structures. *Social networks*, 21(4):375–395, 2000.

[3] S. Borgatti, M. Everett, and L. Freeman. Ucinet for windows: Software for social network analysis. *Harvard Analytic Technologies*, 2006, 2002.

[4] V. Colizza, A. Flammini, M. Serrano, and A. Vespignani. Detecting rich-club ordering in complex networks. *Nature Physics*, 2(2):110–115, 2006.

[5] P. Erdős and A. Rényi. *On the evolution of random graphs*, volume 5. 1960.

[6] M. Gladwell. *The tipping point: How little things can make a big difference*. Little, Brown and Company, 2000.

[7] S. Goel, R. Muhamad, and D. Watts. Social search in small-world experiments. In *World Wide Web*, pages 701–710, 2009.

[8] D. Greer. Incidence of the terror during the french revolution: A statistical interpretation author: Donald greer, publisher: Peter. 1935.

[9] S. Hoory, N. Linial, and A. Wigderson. Expander graphs and their applications. *Bulletin of the American Mathematical Society*, 43(4):439–562, 2006.

[10] M. Humphries. Roman senators and absent emperors in late antiquity. *Acta ad archaeologiam et artium historiam pertinentia*, 17:27–46, 2003.

[11] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *SIGKDD Conference on Knowledge discovery and data mining*, pages 137–146, 2003.

[12] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170, 2000.

[13] G. Kortsarz and D. Peleg. On choosing a dense subgraph. In *FOCS*, pages 692–701, 1993.

[14] P. Krapivsky and S. Redner. Statistics of changes in lead node in connectivity-driven networks. *Phys. Rev. Letters*, 89(25):258703, 2002.

[15] C. Kraus and T. Livius. *Livy, Ab urbe condita*. CUP, 2010.

[16] R. Kumar, P. Raghavan, S. Rajagopalan, D. Sivakumar, A. Tomkins, and E. Upfal. Stochastic models for the web graph. In *Foundations of Computer Science (FOCS), Symposium on*, pages 57–65, 2000.

[17] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *World Wide Web*, pages 591–600, 2010.

[18] S. Lattanzi and D. Sivakumar. Affiliation networks. In *Symposium on Theory of computing(STOC)*, pages 427–434, 2009.

[19] D. Lee, S. Maeng, and J. Lee. Scaling of nestedness in complex networks. *Journal of the Korean Physical Society*, 60(4):648–656, 2012.

[20] J. Leskovec and E. Horvitz. Planetary-scale views on a large instant-messaging network. In *World Wide Web*, pages 915–924, 2008.

[21] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densification and shrinking diameters. *Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):2, 2007.

[22] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney. Statistical properties of community structure in large social and information networks. In *World Wide Web*, pages 695–704, 2008.

[23] N. Linial, D. Peleg, Y. Rabinovich, and M. Saks. Sphere packing and local majorities in graphs. In *Theory and Computing Systems*, pages 141–149. IEEE, 1993.

[24] J. McAuley, L. da Fontoura Costa, and T. Caetano. Rich-club phenomenon across complex network hierarchies. *Applied Physics Letters*, 91:084103, 2007.

[25] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *Workshop on Social Networks (WOSN'08)*, 2008.

[26] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Internet Measurement Conference (IMC'07)*, 2007.

[27] M. Newman. *Networks: an introduction*. Oxford University Press, 2010.

[28] T. Opsahl, V. Colizza, P. Panzarasa, and J. Ramasco. Prominence and control: The weighted rich-club effect. *Physical review letters*, 101(16):168702, 2008.

[29] D. Peleg. Local majorities, coalitions and monopolies in graphs: a review. *Theoretical Computer Science*, 282(2):231–257, 2002.

[30] M. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. *Physical Review E*, 78(2):026101, 2008.

[31] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi. On the Evolution of User Interaction in Facebook. In *Workshop on Social Networks* , 2009.

[32] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications (Structural Analysis in the Social Sciences)*. 1994.

[33] S. Wasserman and K. Faust. *Exploratory Social Network Analysis with Pajek (Structural Analysis in the Social Sciences)*. 2005.

[34] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *nature*, 393(6684):440–442, 1998.

[35] X. Xu, J. Zhang, P. Li, and M. Small. Changing motif distributions in complex networks by manipulating rich-club connections. *Physica A: Statistical Mechanics and its Applications*, 2011.

[36] X. Xu, J. Zhang, and M. Small. Rich-club connectivity dominates assortativity and transitivity of complex networks. *Phys. Review E*, 82(4):046117, 2010.

[37] S. Zhou and R. Mondragón. The rich-club phenomenon in the internet topology. *Communications Letters, IEEE*, 8(3):180–182, 2004.

[38] V. Zlatic, G. Bianconi, A. Diaz-Guilera, D. Garlaschelli, F. Rao, and G. Caldarelli. On the rich-club effect in dense and weighted networks. *European Physical Journal B-Condensed Matter and Complex Systems*, 67(3):271–275, 2009.