



<b>Title</b>	Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks
<b>Authors(s)</b>	Nick, Bobo, Lee, Conrad, Cunningham, Pádraig, Brandes, Ulrik
<b>Publication date</b>	2013-08-28
<b>Publication information</b>	Nick, Bobo, Conrad Lee, Pádraig Cunningham, and Ulrik Brandes. "Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks." IEEE, August 28, 2013. <a href="https://doi.org/10.1145/2492517.2492569">https://doi.org/10.1145/2492517.2492569</a> .
<b>Conference details</b>	2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), 25-28 August 2013, Niagara Falls, Ontario, Canada
<b>Publisher</b>	IEEE
<b>Item record/more information</b>	<a href="http://hdl.handle.net/10197/7544">http://hdl.handle.net/10197/7544</a>
<b>Publisher's statement</b>	© ACM 2013. This is the author's version of the work. It is posted here for your personal use. Not for redistribution. The definitive Version of Record was published in Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, <a href="http://dx.doi.org/10.1145/10.1145/2492517.2492569">http://dx.doi.org/10.1145/10.1145/2492517.2492569</a> .
<b>Publisher's version (DOI)</b>	10.1145/2492517.2492569

Downloaded 2026-05-01 23:49:52

The UCD community has made this article openly available. Please share how this access benefits you. Your story matters! (@ucd\_oa)



© Some rights reserved. For more information

# Simmelian Backbones: Amplifying Hidden Homophily in Facebook Networks

Bobo Nick  
University of Konstanz  
bobo.nick@uni-konstanz.de

Conrad Lee  
University College Dublin  
conradlee@gmail.com

Pádraig Cunningham  
University College Dublin  
padraig.cunningham@ucd.ie

Ulrik Brandes  
University of Konstanz  
ulrik.brandes@uni-konstanz.de

**Abstract**—Empirical social networks are often aggregate proxies for several heterogeneous relations. In online social networks, for instance, interactions related to friendship, kinship, business, interests, and other relationships may all be represented as catchall “friendships.” Because several relations are mingled into one, the resulting networks exhibit relatively high and uniform density. As a consequence, the variation in positional differences and local cohesion may be too small for reliable analysis.

We introduce a method to identify the essential relationships in networks representing social interactions. Our method is based on a novel concept of triadic cohesion that is motivated by Simmel’s concept of membership in social groups. We demonstrate that our *Simmelian backbones* are capable of extracting structure from Facebook interaction networks that makes them easy to visualize and analyze. Since all computations are local, the method can be restricted to partial networks such as ego networks, and scales to big data.

## I. INTRODUCTION

Empirical social networks are often aggregate proxies for several heterogeneous relations, especially when collected from social media and other communication systems. For instance, communication may be triggered by a desire to exchange gossip between close friends or to transmit formal notifications between business partners. When these diverse types of interaction are captured in a single dataset, the overlap of social circles blurs the structure of primary groups that actors are affiliated with.

As an example, consider the network of Facebook friendship relations between students in Figure 1a, where individuals are colored according to the dorm they reside in. While the layout suggests that clusters may be aligned with dorms to some extent, this property is more pronounced in the substructure shown in Figure 1b. We hasten to add that this substructure is derived from the network structure alone, i.e., without taking the dorm attribute into account.

We here propose a method to extract such backbone structures by restricting a network to relationships most strongly embedded in social groups. While there are many conceptions of social groups, “primary” groups are of particular interest:

Over the years sociologists have distinguished various kinds, or what Simmel called “forms” of human groups. Among these, one form in particular has continued to interest investigators for more than a century. Groups that are relatively small, informal, and involve close personal ties – those that Tönnies characterized as based on *Gemeinschaft*, Durkheim

portrayed as reflecting *solidarité organique*, and both Spencer and Cooley described as *primary* – remain at the core of the discipline. [3]

We focus therefore on the identification of strong ties that are embedded in primary groups. While common technical terms include, for instance, clusters and modules we henceforth refer to these primary groups as *communities*.

Note that our goal is not to assign community memberships themselves. As we will argue below, structural embeddedness can be defined locally, without knowing the surrounding communities or even postulating a formal definition for them. The reduction to strong, embedded ties is thus substantially different from common approaches to community detection which assess the relation between internally dense and externally sparse actor groups [4].

In the next section, we motivate our approach by substantive theory on social groups and structural embeddedness. Simmelian backbones are introduced in Section III, and evaluated in Section IV. We conclude with a brief discussion.

## II. BACKGROUND

### A. Tie Dependencies

Community detection algorithms generally equate communities with relatively dense subgroups, where ‘relatively dense’ means that within-group density is higher than expected with respect to some null model or substantially higher than between-group density. Modularity’s null model, for example, is based on a planted partition model, which is a slight generalization of Erdős-Rényi random graphs [5] with different tie probabilities within and between prespecified groups. Within each planted partition, edges are assumed to form independently of each other.

The processes that drive the formation of social networks, however, go beyond the level of dyadic variables. Rather, ties are systematically patterned and embedded in local structures, such as triangles and cycles. The presence of ties is likely to conditionally depend on the presence of other ties. A typical examples are common neighbors. Indeed, such conditional dependencies fundamentally drive network evolution: “Without dependence among ties, there is no emergent network structure.” [6]

The design of valid network models therefore requires careful consideration of such dependencies. For example, Markov random graphs [7], a subset of exponential-family

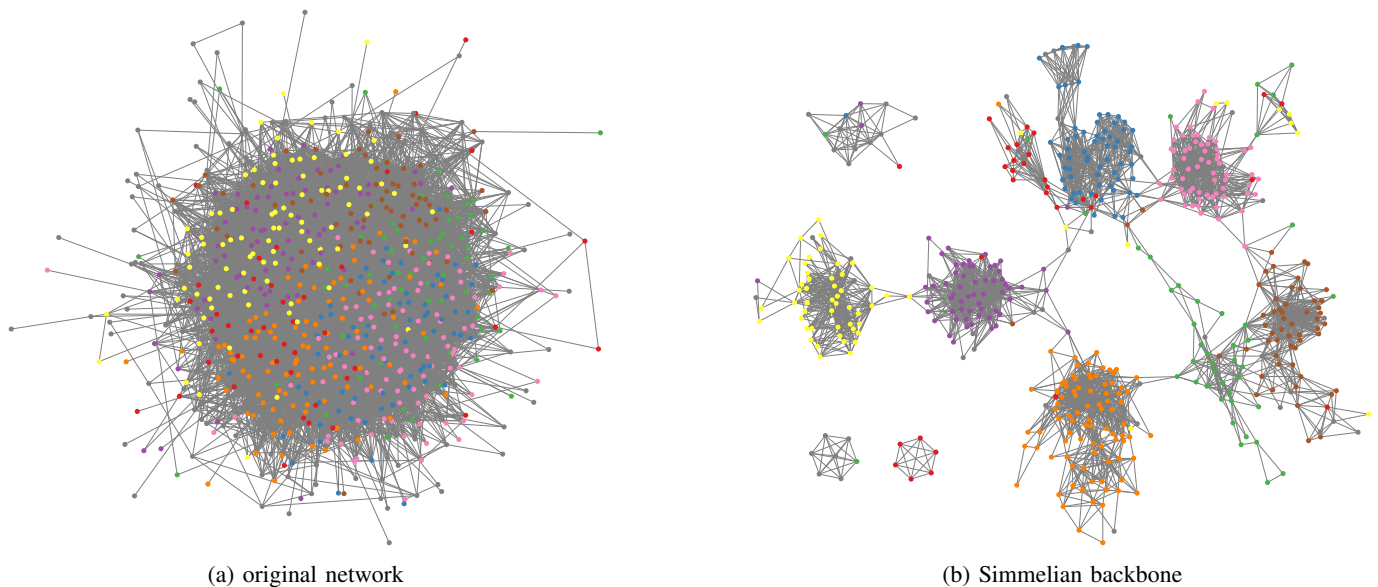


Fig. 1: Facebook friends at California Institute of Technology (Caltech), September 2005. Color encodes dormitory attribute; gray for missing value. (a) giant component as analyzed in [1], [2]. (b) reduced network according to “at least 5 overlapping top-10 ranked neighbors” w.r.t. Simmelian tie strength and after removing edges that are no longer Simmelian in the backbone.

random graph models [8], include conditional dependence of incident ties. It is hence surprising that community detection approaches such as modularity do not seek to exploit such fundamental dependencies; instead, they are usually based on unconditional dyadic aggregate information, such as density and nodal degrees.

While the “cohesive” groups detected by such community detection methods do indeed correspond to relatively dense substructures, they can not be expected to exhibit other distinguished qualities that can be expected by substantive sociological arguments, such as high transitivity.

We illustrate this point in Figure 2, which provides descriptive statistics on the interdependency of dyadic relationships in the Facebook100 data consisting of Facebook friends at collegiate institutions in the U.S. [1], [2]. In each network, the probability of observing a tie conditioned on the number of shared friends increases with the number of shared friends. The number of shared friends equals the number of triangles a dyad can close and is sometimes referred to as its *structural embeddedness*. The effect is roughly linear for small to moderate numbers of common neighbors and then saturates.<sup>1</sup> When the networks are partitioned into density-based communities (using the hierarchical Louvain modularity maximization approach [10]), conditional probabilities of intra-community dyads are lifted (which is consistent with increased density) but the functional form remains largely the same; still, for instance, a small number of common friends significantly increases conditional probabilities of intra-community edges.

<sup>1</sup>For this reason, the specification of transitivity effects in exponential-family random graph models commonly involves geometrically-weighted shared partner statistics [9].

## B. Simmelian Ties and Social Groups

Local density in the sense of modularity is only one of many possible proxies for the cohesiveness of relationships among group members. We now show that sociological theory suggests that structural embeddedness (i.e., the number of triangles in which an edge is embedded, or the number of common neighbors shared by the endpoints of an edge) better accounts for the quantitative and qualitative differences between edges that exist within communities and those that exist between communities. Backed by sociological theory, we thus shift the perspective from an unconditional dyadic view towards a conditional triadic interpretation, and posit that (social) communities are formed by *strong* and highly *redundant* ties between members.

In social networks, dyadic relationships are assumed to depend on the structural environment in which they are embedded [6]. Network formation theories attach particular importance to triadic settings [11], [12]. According to Simmel [13], *triads* (sets of three actors) are fundamentally different from *dyads* (sets of two actors) by way of introducing mediating effects. On the other hand, “the further extension to four or more persons by no means correspondingly modifies the group any further.” [14] Krackhardt goes so far to conclude that “the key to understanding the quality of a tie between two actors can be reduced to asking whether it is part of a strong triad or not” [11] and classifies dyadic relationships accordingly: “super-strong” *Simmelian ties* are embedded in at least one triangle of reciprocated ties, while sole-symmetric and asymmetric ties are not; cf. Figure 3.

Prominent examples of triadic perspectives include Heider’s structural balance theory [15], [16] and Granovetter’s theory of the strength of weak ties [17]. As others have noted, “Granovetter’s argument contends that weak ties provide better connections to different social milieus because they usually

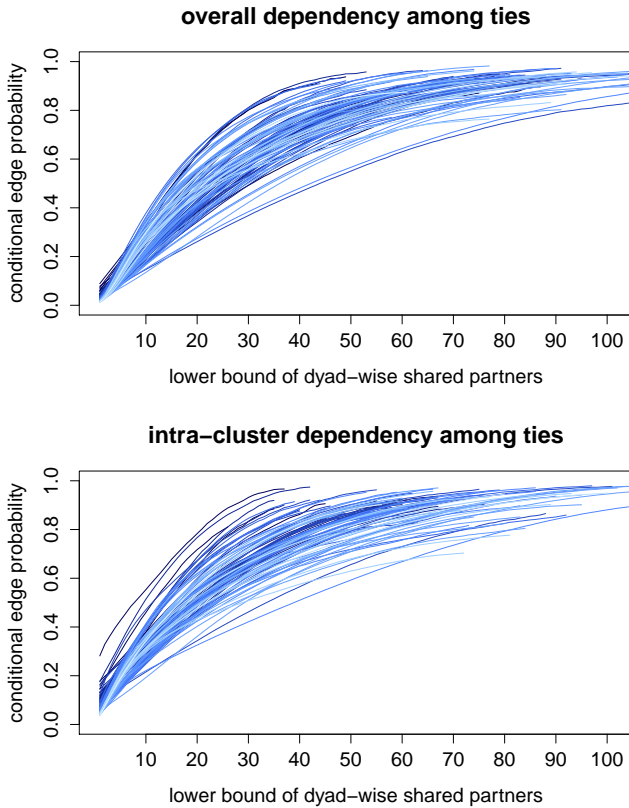


Fig. 2: Conditional edge probabilities in networks of Facebook friends (data from [1], [2]). The  $x$ -axis conditions on a minimum number of triangles a dyad is contained in, and the  $y$ -axis gives the fraction of ties in such dyads. Lines are saturated according to networks density which is decreasing with size since average degree is relatively constant across networks. In the lower figure, dyads are restricted to lie within communities found by Louvain modularity maximization [10].

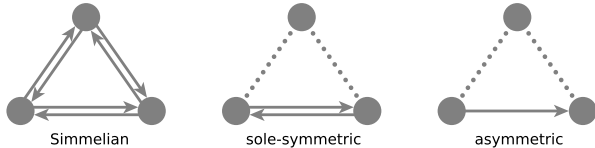


Fig. 3: Qualitatively distinct triadic embeddedness of dyadic relations according to [11]: “super-strong” *Simmelian ties* are embedded in at least one triangle of reciprocated ties, while sole-symmetric and asymmetric ties are not.

connect socially dissimilar people” [18].

Similar to Granovetter, Burt distinguishes between weak ties that provide non-redundant connections to information buried in structural holes, and strong ties, which are much more interconnected (i.e., highly redundant) [19]. A group’s capabilities (and constraints) are then determined by two dimensions: *the redundancy of internal edges* (along the horizontal axis of Figure 4) and *the non-redundancy of external edges* (along the vertical axis). Burt argues that a group achieves maximum performance when its internal edges are strong, and its external edges are weak: “while brokerage across structural holes is the

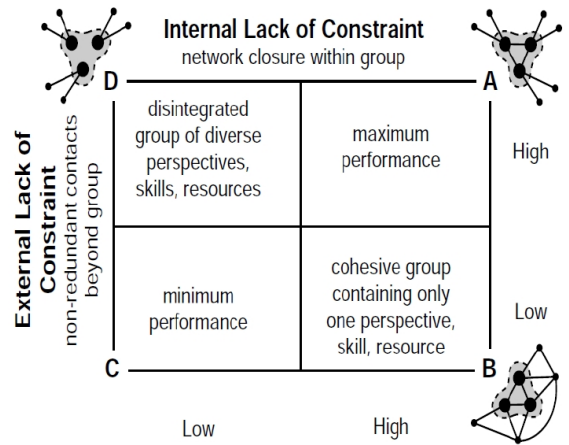


Fig. 4: Group performance according to Burt’s notion of constraint: “while brokerage across structural holes is the source of added value, closure can be critical to realizing the value buried in the structural holes” (reproduced from [20]).

source of added value, closure can be critical to realizing the value buried in the structural holes” [20]; cf. Figure 4.

Our intended notion of triadic cohesion is driven by this conception of a high-performing group with strong internal connections and weak external connections, as in the upper-right quadrant of Figure 4. This notion, which is fundamentally based on the embeddedness of edges, is thus very different from aggregate density-based measures such as modularity, which do not consider embeddedness at all. Against this background, we now specify a method for distinguishing a) intra-cluster edges: “strong”, Simmelian, structurally embedded ties exhibiting a high level of redundancy and homogeneity, from b) inter-cluster edges: “weak”, non-Simmelian, less embedded ties that are rather diverse and heterogeneous.

### III. SIMMELIAN BACKBONES

In the following we propose a framework to extract a substructure consisting of ties that are strong and redundant. The approach is based on a triadic notion of cohesion and consists of the following steps:

- A. If input tie strength is uniform, assign to ties the number of triangles they are embedded in (Simmelian strength).
- B. For each actor, rank all alters by tie strength.
- C. For each (strong) tie, determine its redundancy.
- D. Filter ties that are weak or not redundant.

#### A. Tie Strength

Our procedure requires that each edge is labeled with a preliminary weight that is a proxy for social importance. While such weights are rarely explicitly known, common proxies are frequency of contact in communication networks or shared attributes in recommender systems. If such a weight is available, then one can simply use this and proceed to B.

Even if there is no explicit information on edge weight, as is the case with the Facebook data examined here, sociologically informed strengths can be extracted from the network

structure itself. According to Granovetter, structural embeddedness is an important proxy for tie strength and in many substantive applications embeddedness can be characterized by triadic configurations.

While the specific weighting technique will depend on what ties represent, a generic proxy for uniformly weighted friendship networks is Simmelianess [11]. Dekker [21] lists several tie-level measures of Simmelianess and claims that “The main advantage of the Simmelian tie strength measures is that they are firmly rooted in a substantive theory.” Under the binary variant, a tie is considered Simmelian if it is part of a triangle (cf. Figure 3). A straightforward extension to measure the degree of Simmelianess for an edge  $\{i, j\}$  is to count the number of neighbors shared by  $i$  and  $j$  (i.e., the number of distinct triangles a tie is embedded in). Dekker argues that shared partners “are central in opinion formation about behavior and intentions” and ever more shared partner impute constraints “on each others behavior, because more of their behaviors become observable to each other.” [21]

For the purpose of this paper, and the examples in the next section, we use this weighting technique (i.e., the number of shared neighbors) as a preliminary proxy for social importance in unweighted social networks. We discuss the computational complexity of this approach below.

## B. Ranking Ties

A pivotal factor in our approach is the following: In correspondence with Burt’s notion of constraint, we do not consider a global, network-level ranking of ties but advocate a local assessment on the level of actor neighborhoods. Let ego be an actor in a network with ordinal tie weights. It does not matter whether the relation is directed or not. We define ego’s *rank-ordered neighborhood* as the list of adjacent actors (alters) sorted from strongest to weakest tie between ego and alter.

$N_{ego}^{\rightarrow}$  : ego’s alters ranked from strongly to weakly tied

This actor-level ranking procedure is very different from ranking all ties globally, for at least two reasons. First, the procedure introduces *asymmetry* between connected pairs of nodes, even in undirected networks: for example, whereas B may only rank eleventh in A’s list, A may rank third in B’s list. Secondly, the procedure can deal with *heterogeneity*: edges which have the same weight from a global perspective may have very different rankings from a local perspective. For example, a tie embedded in ten triangles in a very dense region of the graph may rank low, whereas an equally embedded edge in a sparser portion of the graph may rank high. Whether the input was directed or not, the result of this step is a (rank-weighted) directed graph which we refer to as the *ranked neighborhood graph*.

## C. Tie Redundancy (Structural Embeddedness)

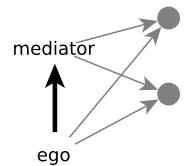
There are various ways of filtering the edges of a weighted graph based on derived weights and absolute or relative thresholding (see, e.g., [22], [23], [24]). Many of these, however, bear little relationship with the Simmelian idea of groups.

Based on the ranked neighborhood graph, we therefore introduce a novel measure of triadic cohesion. The crucial idea is to compare the local perspectives of ego and alter on the social importance of others: a strong (i.e., highly ranked) tie is considered strongly embedded,<sup>2</sup> if its endpoints have similar views on highly ranked neighbors — that is, the actors incident to a *triadic cohesive* tie are relatively strongly connected to each other and also relatively strongly connected to relatively many common neighbors (and thus embedded in very strong Simmelian groups).

Formal definitions of the degree of embeddedness thereby correspond to a similarity assessment of ranked neighborhood lists. In this way, we can avoid that (additional) weak ties to distinct neighbors degrade the evaluation of top-ranked (strong) redundancies. Two examples, one parametric and one non-parametric, follow below.

1) *Parametric variant*: A granularity parameter  $k$  specifies the rank after which ties are not considered to be relatively strong anymore. The degree to which a strong tie is structurally embedded is then defined as the *overlap* of common neighbors among the top- $k$  entries of  $N_{ego}^{\rightarrow}$  and  $N_{alter}^{\rightarrow}$ .

This procedure corresponds to evaluating ego’s (strong) connections in the directed top- $k$  ranked neighborhood graph for mediating positions in transitive triplets; see adjacent diagram. Transitive (mediated) triplets and balanced neighborhoods have proven crucial, e.g., in modeling the evolution of longitudinal networks [25].



A general reference for comparison of top- $k$  lists is [26].

2) *Non-parametric variant*: While a granularity parameter  $k$  appears convenient for exploring structural embeddedness at different levels, it may also prove to be a burden in the absence of substantive expectations on the degree of embeddedness.

A non-parametric alternative is to automatically determine, for each edge, a  $k$  that maximizes the overlap among the top- $k$  entries of  $N_{ego}^{\rightarrow}$  and  $N_{alter}^{\rightarrow}$  as above. Because this involves prefixes of different length, however, the number of common entries is normalized by the size of the union of entries. In other words, we compute the Jaccard coefficient for prefixes of any length and pick the maximum value. Note that the classical Jaccard coefficient (i.e. the comparison of all neighbors) only provides a lower bound for the best prefix Jaccard coefficient.

While the non-parametric variant allows for even more heterogeneity across edges, we do not yet have a good grasp of the consequences because of its more difficult interpretation.

We note in passing that both variants can take into account reciprocity in the redundancy assessment by identifying alter’s rank in  $N_{ego}^{\rightarrow}$  with ego’s rank in  $N_{alter}^{\rightarrow}$ .

## D. Filtering Ties

We define the *Simmelian backbone* of a social network by applying a (global) threshold on the derived (local) measure of triadic cohesion. Minimum requirements include, e.g.,

<sup>2</sup>In the present setting we find the term (structural) embeddedness more intuitive than redundancy.

- a minimum number of common top- $k$  neighbors, or
- a required prefix Jaccard coefficient.

In other words, we filter the edges of a ranked neighborhood graph first by their relative strength (i.e., their rank in  $N_{ego}^{\rightarrow}$ ) and then, additionally, by their strong embeddedness (i.e., the similarity of  $N_{ego}^{\rightarrow}$  and  $N_{alter}^{\rightarrow}$  in their top ranks).

In the so reduced network, remaining ties are directly and indirectly attached to top-ranked neighbors: we expect to observe high similarity between the ties of ego and the ties of those alters to whom ego still has a tie after the filtering (while weak and less embedded connections are implicitly filtered away). Based on this transformation, therefore, we induce an individual-level attachment to groups corresponding to Burt’s notion of constraint. Indeed, the proposed heuristic can be interpreted as a pairwise comparison of ego networks, since all calculations are solely based on local information.

As we will demonstrate in Section IV, social communities and bridges between them become rather obvious in the filtered network. For convenience, inter-community bridges can be filtered by removing also edges that are no longer Simmelian in the backbone. However, where appropriate, this should not be done for edges in the 1-core that are not in the 2-core, since these connect otherwise isolate actors with the community to which they are most attached.

If the communities themselves are of interest, any technique from the broad range of existing community detection algorithms can be applied to partition the network via its backbone. As illustrated in the next section, the balance between inter- and intra-community edges has altered strongly in favor of the latter, so that community detection is easy.

We conclude this section with a word on the computational complexity of this approach, including the weighting scheme. The triangles of a graph  $G$  can be listed in time  $\mathcal{O}(m\alpha(G))$  [27], where  $m$  is the number of edges and  $\alpha(G) \leq \sqrt{m}$  its arboricity. The arboricity is defined as the minimum number of forests needed to cover the edges. Since it is bounded from above by the  $h$ -index of the graph and the latter has been found to be small for networks of social relations [28], the arboricity can be expected to be small as well. Both Simmelian tie strength and the degree of embeddedness can be determined by a slight variation of the triangle listing algorithm. In the non-parametric variant the latter requires  $\mathcal{O}(m\alpha(G))$ , again, and in the parametric one it becomes linear for fixed  $k$ . Moreover, neighborhoods can be rank-ordered in linear time using bucket sort, since tie weights are bounded by  $n - 2$ , the maximum number of shared neighbors in a graph with  $n$  nodes. Overall, the ranked neighborhood graph of an unweighted directed or undirected graph (including the weighting technique described above) is constructed in  $\mathcal{O}(m\alpha(G))$  time where  $\alpha(G)$  is expected to be small.

#### IV. EVALUATION

We illustrate a potential use case of Simmelian backbones by re-analyzing the 100 networks of Facebook friendships in the Facebook100 dataset. These networks were introduced and analyzed in [1], [2]; they range in size from 769 nodes and 17k edges to 36k nodes and 1.6m edges. The data has

many desirable characteristics; for example, it comes directly from Facebook and is not sampled, and it comes from a service which at the time of data collection was widely and intensively used by students. Furthermore, the dataset includes node information on several attributes, derived from the Facebook profiles: gender, year of graduation, dormitory, primary academic major, secondary academic major, high school, and student/faculty status. Not every user filled in every field of the profile, and so for all attributes many values are missing (except for student/faculty status, for which no attribute values are missing).

Traud et al. used a modularity maximization technique to detect communities in their investigation of these networks, and found that in some cases dorms displayed a high level of homophily within communities, and in other cases the year of graduation appeared to be the attribute most related to community structure. In particular, the dorm attribute tended to have high homophily in communities found at small institutions, whereas the year attribute was more enriched in the communities detected at larger institutions. Findings in [29] indicate that this tendency may be an artifact caused by limitations of modularity-based community detection, which is known to suffer from a resolution limit [30].<sup>3</sup> Thus, despite previous work on the topic, the relationship between the community structure, and the level of homophily for each of the node attributes, remains somewhat uncertain.

##### A. Visual Exploration

Visualization of these networks has proven difficult. Three of the Facebook100 datasets were used for the Visual Analysis of Complex Networks (VACN) challenge.<sup>4</sup> While some interesting visualizations emerged, none of them clearly displayed both the connections between and within communities.

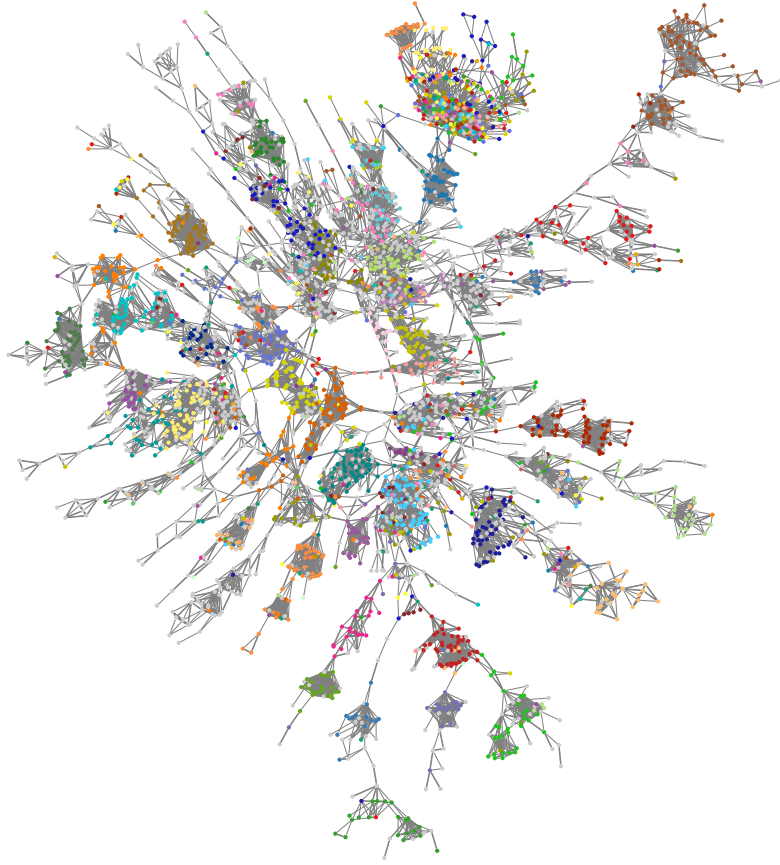
We provide anecdotal evidence that extracting the Simmelian backbone and visualizing the resulting subgraph yields more informative visualizations of the community structure present in Facebook100 networks. In addition to the Simmelian backbone of the Caltech network visualized in Figure 1b, we show the backbone of the University of Chicago network in Figure 5.

The Caltech network received extensive discussion in [1], in which the authors note that at Caltech the residential house system is particularly important for the formation of social relationships. However, the network visualization of the Caltech network presented in that paper does not provide a clear picture of how the dormitories are closely related to social life; this relationship is much clearer in Figure 1b.

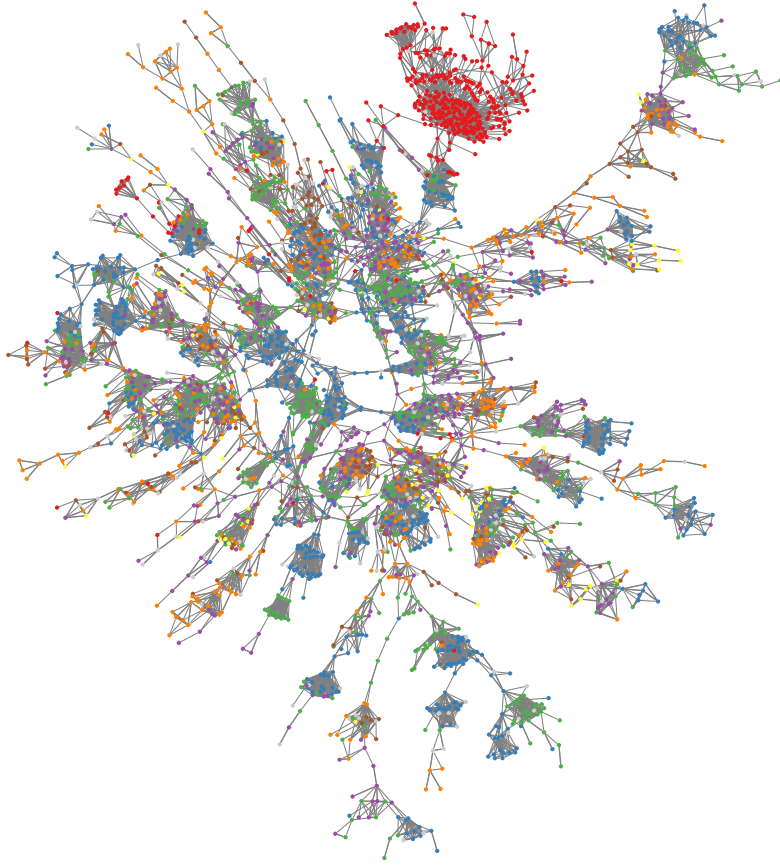
As one of the authors has first-hand experience of the social life there and is included in that dataset, we will pay particular attention to the Simmelian backbone of the University of Chicago network, which includes 6,591 nodes and 208,103

<sup>3</sup>For example, maximizing modularity in many cases produced unreasonably large and few communities. A parameterized version of modularity was shown to alleviate these problems, but Traud et al. did not use the parameterized definition in their investigation.

<sup>4</sup>With participation from tools Gephi, i2’s Analyst Notebook, Pajek, Tulip, and visone. See <http://sociograph.blogspot.com/2011/02/visualizing-large-facebook-friendship.html> for details including the two best visualizations of the Caltech network – compare these to Figure 1b.



(a) Nodes colored by dorm



(b) Nodes colored by year of graduation

Fig. 5: Giant component of Simmelian backbone of Facebook friends at the University of Chicago (3854 nodes, 19724 edges): (a) Nodes colored according to dorm they live in. Because there are dozens of distinct values for this attribute, not all colors are visibly distinct. (b) Nodes colored according to year of graduation. Note the new arrivals on the right (red colored). Grey nodes indicate missing attribute values.

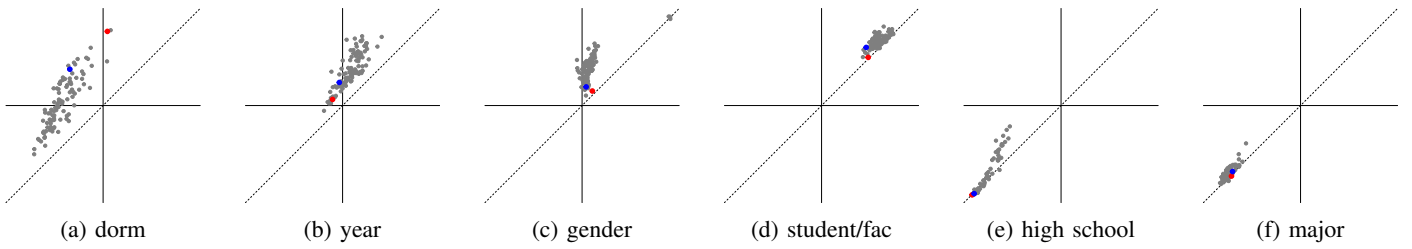


Fig. 6: Share of ties linking nodes with the same attribute value in original network ( $x$ -axis) and Simmelian backbone ( $y$ -axis). Each dot represents one network of Facebook friends with Caltech (red) and University of Chicago (blue) highlighted. For networks above the diagonal, homophily is stronger in the backbone.

undirected edges. The residential “houses” are known to be of utmost importance to the social life at the University of Chicago: upon entering the university, every student is required to spend at least one year living in the dormitory system, and the friendships formed in this stage of college—for many students, the first time living away from home—often endure for years. According to the University of Chicago website, “each house represents a tight-knit community of students, resident faculty masters, and residential staff, who live, relax, study, dine together at House Tables, engage, socialize, and learn from each other.”<sup>5 6</sup>

As we can see by zooming in closely on Figure 5a, in which nodes are colored by their house, the Simmelian backbone does a good job of recognizing clustered structures in which the dorm attribute displays a high level of homophily, providing a strong validation of the approach. The major exception to this trend is the large cluster located at the 3:30 position. In Figure 5b we see that all of these nodes share the same year; in fact, the metadata indicates that these red colored nodes have 2009 as their graduation year. This dataset was collected in September 2005. In the middle of that month, the University of Chicago class of 2009 moved into their dorms and attended orientation week, during which they interacted with each other and had very little interaction with students graduating in other years, most of whom would not arrive back in Chicago until the end of September when the Fall term started. It is interesting to see the first-year students isolated in their own group, having not yet met and integrated with the older members of their dorms. We note that this feature is not apparent in visualizations which include all edges in the graph—we first noticed it when looking at the Simmelian backbone.

## B. Quantitative Description

We persistently observe the same effect for almost all network and all relevant attributes, namely that backbone edges

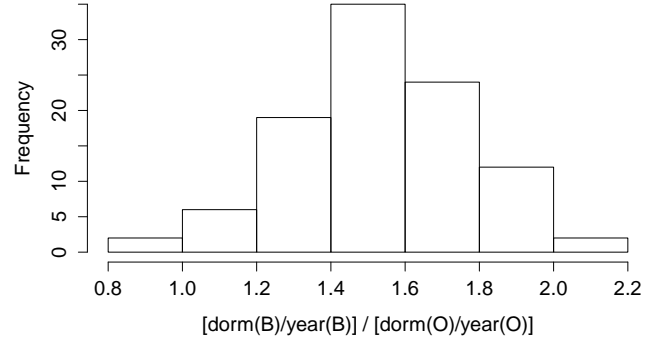


Fig. 7: The prominence of dorm-homophily over year-homophily increases on average by a factor of 1.5 in the Simmelian backbones (B) when compared to the original networks (O).

display a stronger tendency for homophily than the entire set of edges. The share of homophily in the original network and in the Simmelian backbone for each Facebook100 network is contrasted in Figure 6.

Each point in the scatterplots represents one institution in the Facebook100 data. Points above the dotted line through the origin indicate that homophily in the Simmelian backbone is increased compared to the original network with respect to a focal attribute. In detail, for each attribute, edges are classified as follows: “homophily edges” connect students having the same attribute value, while “heterophily edges” connect students having a different attribute value. Then, the share of “homophily edges”,

$$\frac{\#\text{homophily edges}}{\#\text{homophily edges} + \#\text{heterophily edges}},$$

in the original network ( $x$ -axis) is contrasted with the same fraction of edges in the Simmelian backbone ( $y$ -axis); centered at  $(0.5, 0.5)$ . Edges with attached student for which the focal attribute is undefined are not taken into account in this ratio calculation.

The scatterplots in Figure 7 indicate which attributes are most important factors in network formation on the aggregate level: dormitory, graduation year, gender, and student/faculty status. Our Simmelian backbones reveal that homophily in the first three attributes is much stronger than can be noticed from the original networks. In addition, they also show that the relative importance of attributes is shifted after denoising

<sup>5</sup><http://csl.uchicago.edu/feature/announcing-new-residence-hall-and-dining-commons>

<sup>6</sup>The meta-data does not explicitly indicate that the “dorm” meta-data represents the housing system at the University of Chicago, but this is strongly suggested by the data itself, as the number of distinct values corresponds quite closely to the number of houses and fraternity houses. At the time of data collection, the housing system consisted of 10 residence halls (physical buildings) which were further subdivided into 37 houses, which typically represent a physically adjacent wing of a residence hall. House sizes range from 100 to 37 students, with an average size of 70. Furthermore, there were somewhere between ten and twenty fraternity houses. Thus some of the dorm meta-data may indicate fraternity membership.

the networks. Figure 7 in particular suggests that dorm-homophily, which could previously only be observed in the smaller networks [1], [2], was subdued by year-homophily.

These numbers together with the above network visualizations clearly suggest that within-dormitory communities are more prominent than could previously be detected.

## V. DISCUSSION

We presented an approach to uncover deeply embedded relations that are likely to be internal to primary groups, referred to as communities. The method serves, among other things, to reveal essential structure in networks with insufficient variation in local density for standard community detection approaches to work.

The main ingredient of our method is the concept of *ranked* neighborhood comparisons to infer the structural embeddedness of ties. The corresponding notion of triadic cohesion has three desirable aspects: (i) it implements sociological theory and should therefore be easier to justify in empirical settings, (ii) it is defined locally and therefore applicable to ego networks as well, and (iii) it is computationally efficient and scales to big network data.

By way of re-analysis, we showed that Simmelian backbones of Facebook friendship networks may serve to reveal primary groups that provided strong evidence of homophily and had been difficult to detect previously.

Although our examples are all based on static undirected unweighted networks, the method can be adapted straightforwardly to directed, weighted, and temporal network data. In fact, the social vector clocks of [31] yield networks with tie strength that align well with the underlying theory.

There are several other directions for future research. We are currently exploring alternative definitions of relatively strong structural embeddedness and different filtering schemes. More importantly, we are interested in structural properties of Simmelian backbones related to invariants of the input graph as well as their impact on community detection algorithms.

## ACKNOWLEDGMENT

We thank Amanda L. Traud, Peter J. Mucha, Mason A. Porter, Eric D. Kelsic, and Adam D'Angelo for providing the Facebook100 dataset. This work is supported in part by Deutsche Forschungsgemeinschaft under grant Br 2158/6-1, Science Foundation Ireland under grant no. 08/SRC/I1407 (Cliques: Graph and Network Analysis Cluster), and the University of Konstanz under grant FP 665/10.

A free network analysis tool offering Simmelian backbone extraction is **visone**. See [www.visone.info](http://www.visone.info) and its description in the visone wiki.

## REFERENCES

- [1] A. L. Traud, E. D. Kelsic, P. J. Mucha, and M. A. Porter, "Comparing community structure to characteristics in online collegiate social networks," *SIAM Review*, vol. 53, no. 3, pp. 526–543, 2011.
- [2] A. L. Traud, P. J. Mucha, and M. A. Porter, "Social structure of facebook network," *Physica A*, vol. 391, no. 16, pp. 4165–4180, 2012.
- [3] L. C. Freeman, "The sociological concept of 'group': An empirical test of two models," *American Journal of Sociology*, pp. 152–166, 1992.
- [4] S. Fortunato, "Community detection in graphs," *Physics Reports*, vol. 486, pp. 75–174, 2010.
- [5] P. Erdős and A. Rényi, "On random graphs," *Publicationes Mathematicae Debrecen*, vol. 6, pp. 290–297, 1959.
- [6] U. Brandes, G. Robins, A. McCranie, and S. Wasserman, "What is network science?" *Network Science*, vol. 1, no. 1, pp. 1–15, 2013.
- [7] O. Frank and D. Strauss, "Markov graphs," *Journal of the American Statistical Association*, vol. 81, pp. 832–842, 1986.
- [8] G. Robins, P. Pattison, Y. Kalish, and D. Lusher, "An introduction to exponential random graph (p\*) models for social networks," *Social Networks*, vol. 29, no. 2, pp. 173–191, 2007.
- [9] T. A. B. Snijders, P. E. Pattison, G. L. Robins, and M. S. Handcock, "New specifications for exponential random graph models," *Sociological Methodology*, vol. 36, no. 1, pp. 99–153, 2006.
- [10] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2008, no. 10, p. P1000, 2008.
- [11] D. Krackhardt, "Simmelian ties: Super strong and sticky," in *Power and Influence in Organizations*, R. Kramer and M. Neale, Eds. SAGE, 1998, pp. 21–38.
- [12] —, "The ties that torture: Simmelian tie analysis in organizations," *Research in the Sociology of Organizations*, vol. 16, pp. 183–210, 1999.
- [13] G. Simmel, *The Sociology of Georg Simmel*, K. H. Wolff, Ed. Free Press, 1950.
- [14] —, "Individual and society," in *The Sociology of Georg Simmel*, K. H. Wolff, Ed. Free Press, 1950.
- [15] F. Heider, "Attitudes and cognitive organization," *Journal of Psychology*, vol. 21, pp. 107–112, 1946.
- [16] D. Krackhardt and M. S. Handcock, "Heider vs Simmel: Emergent features in dynamic structures," in *ICML 2006*, ser. LNCS, E. Airoldia, D. M. Blei, S. E. Fienberg, A. Goldenberg, E. P. Xing, and A. X. Zheng, Eds., 2007, vol. 4503, pp. 14–27.
- [17] M. Granovetter, "The strength of weak ties," *The American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
- [18] B. Wellman and M. Gulia, "The network basis of social support: A network is more than the sum of its ties," in *Networks in a Global Village. Life in Contemporary Communities*, B. Wellman, Ed. Westview Press, 1999, pp. 83–117.
- [19] R. S. Burt, *Structural holes: The social structure of competition*. Harvard University Press, 1992.
- [20] —, "Structural holes versus network closure as social capital," in *Social Capital: Theory and Research*. De Gruyter, 2001, pp. 31–56.
- [21] D. Dekker, "Measures of Simmelian tie strength, Simmelian brokerage, and the Simmelianly brokered," *Journal of Social Structure*, vol. 7, no. 1, 2006.
- [22] D. S. Goldberg and F. P. Roth, "Assessing experimentally derived interactions in a small world," *PNAS*, vol. 100, no. 8, pp. 4372–4376, 2003.
- [23] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi, "Defining and identifying communities in networks," *PNAS*, vol. 101, no. 9, pp. 2658–2663, 2004.
- [24] G. Melançon and A. Sallaberry, "Edge metrics for visual graph analytics: A comparative study," in *Proc. 12th Intl. Conf. Information Visualization (IV'08)*, 2008, pp. 610–615.
- [25] R. M. Ripley, T. A. Snijders, and P. Preciado, "Manual for RSiena," 2012.
- [26] R. Fagin, R. Kumar, and D. Sivakumar, "Comparing top k lists," in *Proc. 14th Ann. ACM-SIAM Symp. Discrete Algorithms*, 2003, pp. 28–36.
- [27] N. Chiba and T. Nishizeki, "Arboricity and subgraph listing algorithms," *SIAM Journal on Computing*, vol. 14, no. 1, pp. 210–223, 1985.
- [28] D. Eppstein and E. S. Spiro, "The h-index of a graph and its application to dynamic subgraph statistics," in *Algorithms and Data Structures*, ser. Lecture Notes in Computer Science, F. Dehne, M. Gavrilova, J.-R. Sack, and C. D. Tóth, Eds., 2009, vol. 5664, pp. 278–289.
- [29] C. Lee and P. Cunningham, "Benchmarking community detection methods on social media data," *arXiv preprint arXiv:1302.0739*, 2013.
- [30] S. Fortunato and M. Barthelemy, "Resolution limit in community detection," *PNAS*, vol. 104, no. 1, pp. 36–41, 2007.
- [31] C. Lee, B. Nick, U. Brandes, and P. Cunningham, "Link prediction with social vector clocks," *Proc. 19th ACM SIGKDD Conf. on Knowledge, Discovery, and Data Mining*, 2013, to appear.