

---

# Community Finding: Partitioning Considered Harmful

---

**Fergal Reid**

Clique Research Cluster  
University College Dublin  
Dublin, Ireland  
fergal.reid@gmail.com

**Aaron McDaid**

Clique Research Cluster  
University College Dublin  
Dublin, Ireland

**Neil Hurley**

Clique Research Cluster  
University College Dublin  
Dublin, Ireland

## Abstract

Considering a clique as a conservative definition of community structure, we examine how partitioning algorithms interact with cliques. We show that on a wide range of empirical networks, from different domains, significant numbers of cliques are split across different partitions by popular algorithms. We examine the largest connected component of the subgraph formed by retaining only edges in cliques, and apply partitioning strategies that explicitly minimise the number of cliques split. We conclude that, due to the connectedness of many networks, any partitioning community finding algorithm must fail to return at least some significant structure. Moreover, contrary to traditional intuition, strong ties and cliques frequently do cross community boundaries.

## 1 Introduction

Groups of interacting entities can be considered as a complex system. It is popular to examine such systems in terms of the networks their component entities form, to gain insight into properties of the system as a whole. For example, the speed with which a contagion can spread through a system is partly determined by the topology of its underlying network. The way subgroups of entities interconnect is also important to investigate whether useful higher level abstractions – above the level of individual entities – exist in the systems we study. To find such structures, an extensive variety of algorithms have been developed, which attempt to find groups of nodes in the network that are structurally significant in some way; these groups are referred to in the literature as ‘communities’. See [8] for an extensive review of these algorithms, which we will refer to as *Community Finding Algorithms*, or *CFAs*.

CFAs have been put to a range of applications, across several domains. As CFAs are applied ever more broadly, it is important that the structures they find, and the consequences of the design choices that define them are well understood. Particular CFAs should not be assumed to work across all complex networks, merely because they have evaluated well on some. In this research, we argue that certain algorithms, notably CFAs that produce *partitions* of the original network, return incomplete lists of the significant community structure, for at least some empirical networks. We show that certain networks do not lend themselves well to partitioning, and caution against using partitioning algorithms as universal community finding tools.

Each CFA finds structure that corresponds to a particular intuition of what a ‘community’ is; however there is little agreement on how exactly to define community. One idea that is common is that a community should have a high density of edges among its nodes, where density refers to the ratio of actual edges between the nodes of the community to the maximum possible. The bound of this definition is the graph theoretic structure known as a ‘clique’ – a fully connected subgraph, where each node is connected to every other. Cliques, as discussed by Luce et al.[16], have long been

considered as community structure in human social networks. In the domain of social networks, this is very intuitive; if a user is friends with several others on Facebook, all of whom are also friends, then this is a significant structure of common friends. In addition to this intuitive appeal, cliques are rare structures in the networks we study – due to the strict requirement for each node to connect to every other, clique structure is unlikely to arise by chance in a sparse network. Cliques are thus important structures – however, the definition of clique as a community is strict and conservative, for if even one connection is missing – perhaps due to an incompletely observed network – the found community will shrink. Many CFAs thus try and find communities which comprise groups of nodes which are highly connected, but less connected than perfect cliques. However, we posit that a clique is a good conservative lower bound estimate on community structure – a maximally interconnected group of nodes, in a sparse network, always represent some interesting structure.

Many leading CFAs assign communities by partitioning the network, that is, grouping the nodes into *disjoint* subsets, such that each node to one, and only one, subset. This partitioning approach to community finding has become popular, perhaps due to the appeal of treating a complex network as a graph, and the body of literature on graph partitioning problems. Early applications of graph partitioning, such as [11], discuss problems that explicitly require partitions, such as electronic component layout. In this work we are concerned about the completeness of the lists of found community structure on more complex domains. Regarding cliques as underestimates of community structure, we believe that regardless of what specific structures a given CFA finds, to be thorough, it should find, for each clique, *at least* one structure which is a superset of that clique. A CFA – considered as a tool that reveals structure in a complex network – that returns no community in which a group of fully connected nodes are assigned together, is neglecting to provide a complete list of the structures in the network. This is especially true if the clique is large in size.

We show that on many complex networks, partitioning CFAs split cliques, and hence fail to find complete lists of structure. We examine why this occurs, investigating the intuition underlying many partitioning CFAs, and their relationship with cliques. We show, using cliques as a tool, that some traditional intuition about communities is not always correct. Instead, many of the graphs we study exhibit a structure that can be better explained as the ‘pervasive overlap’ discussed in [1],[7] than as independent, weakly-connected modules. We analyse cliques, rather than any new definition from the ‘overlapping community finding’ literature, because we require a definition of structure that is a fundamental and convincing conservative underestimate of community structure. For example, analysing structures such as the percolated  $k$ -cliques of Palla et al.[18] we find no universal  $k$  consistent across networks, with which to evaluate partitioning. We require, for every community, a conservative subset of that community. Rather than choosing a new definition of community and discussing whether it is sufficiently conservative, we instead return to the fundamental definition of the clique alone, and examine its implications in detail. We analyse some of the same data as Leskovec et al.[15] – however, while that influential work sought to investigate the quality of the best community structure, at each scale, by evaluating it in terms of *conductance*, we investigate network structure from a different angle, by using cliques to conservatively estimate community cores, and characterise to what level every one of these may be recovered after partitioning methods are applied, thus considering structure globally across the network.

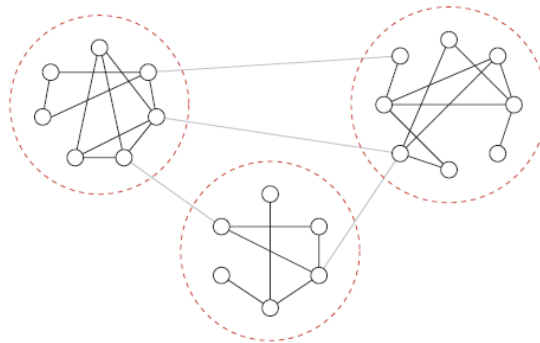


Figure 1: Networks and communities, diagram from Newman[17], illustrative of traditional community intuition

## 2 Experiments

An illustration of the intuition behind many CFAs can be seen in Figure 1 from the influential paper by Newman [17], with well defined separate modules connected by only narrow bridges. This same intuition, conceptualising communities as connected by narrow bridges, can be traced back to the seminal work of Granovetter [9] where he writes: *“If the motivation to spread the rumor is dampened a bit on each wave of retelling, then the rumor moving through strong ties is much more likely to be limited to a few cliques than that going via weak ones; bridges will not be crossed.”* Here, Granovetter is using ‘clique’ in the sociological sense, closer to the modern idea of community, and the key idea is that bridges – narrow connecting links – need to be crossed to carry information between such cliques. This idea is further summed up in the modern review of Fortunato [8] as: *“If it were possible for a clique to move on a graph, in some way, it would probably get trapped inside its original community, as it could not cross the bottleneck formed by the inter-community edges.”* We now show that while this intuition may be appropriate in many cases, the structure of empirical networks does indeed lead to cliques crossing the ‘bottleneck’ formed by inter-community edges.

We conducted experiments to investigate the extent to which commonly used partitioning methods would split the cliques present on empirical network datasets. To keep the number of cliques we consider tractable, and in keeping with the sociological definition of clique [16], we constrain our analysis to *maximal* cliques – for convenience, we will refer to the maximal cliques as simply cliques. We generate the complete list of cliques present in each dataset using the fast Bron-Kerbosch algorithm [3]. We then use the partitioning method under evaluation to assign each node to a community, and characterise how the cliques interact with the partitions found, by examining each maximal clique in turn, checking whether it is fully contained within a partition, or to what extent it has been split across partitions. We quantify and present this for each network, using two distinct partitioning methods; one popular and efficient modularity optimization method [2] and one min-cut optimizing method [6].

### 2.1 Network Datasets Examined

To analyse data from a wide variety of networks, we used network datasets from the SNAP project<sup>1</sup> [15]. These networks provide examples from many different domains, including web2.0, co-citation and product recommendation networks. We also examined several Facebook university social network datasets [19]. The Facebook networks vary in size – we chose to run our experiments on the smaller datasets, due to the computational cost of calculating all maximal cliques. To broaden our analysis, we generated several random snowball samples from a full Twitter network dataset [12]. We also used mobile telecoms data provided by an industrial partner<sup>2</sup>, comprising of the voice call and SMS interactions of a mobile telecoms operator. Like Twitter, this is a large network; we generated 3 random snowball samples of this data to produce tractable network datasets. Finally, we consider the CYC Protein-Protein interaction (PPI) network.

### 2.2 Partition by modularity maximisation

Many of the most popular CFAs are based on the modularity maximization approach of Newman [17]. The modularity function measures community quality as a count of internal edges, less the expected number in a random graph with the same node degrees. Modularity maximization algorithms, such as the fast method by Blondel et al. [2] which we evaluate here, optimise the number of partitions as well as the associated partitioning. While traditional intuition holds that even triangles, or ‘strong ties’, should not cross community boundaries, we are interested in more significant cliques – so we restrict our analysis to cliques of *at least* size 4. We also use a conservative definition of when a clique is ‘split’ – we say a clique is “split at level  $\alpha$ ” if no partition contains more than  $(100 \times \alpha)\%$  of its nodes. We quantify the proportion of cliques that are split by the partitioning of each network in two ways. First, we examine the proportion of cliques of at least size 4 that are split at level  $\alpha = 0.9$ ; Table 1 shows the significant proportions of cliques split at this level. We would have expected, from traditional intuition, that such structures would be contained in the center of the

---

<sup>1</sup><http://snap.stanford.edu/data/>

<sup>2</sup>Idiro Technologies

found communities – not spanning them, and not split by partitions that define found communities. We provide an illustrative example of this effect in Figure 2.

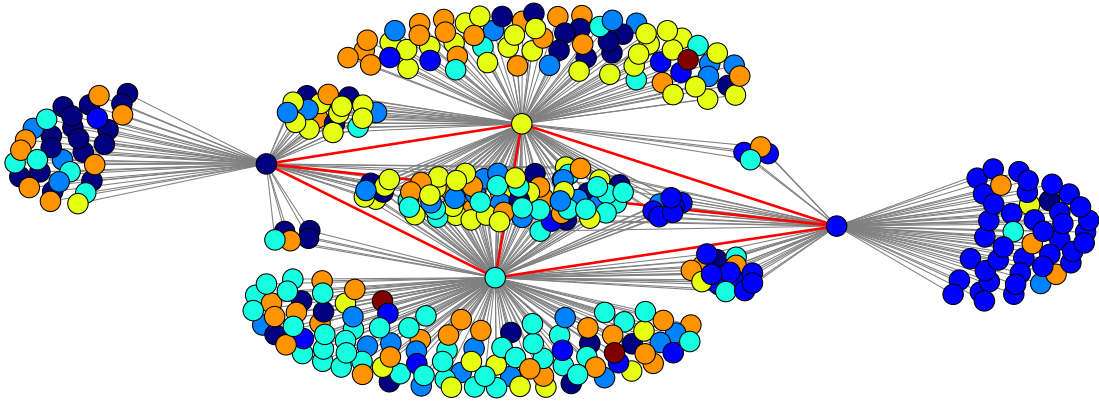


Figure 2: Visualisation of one of the split 4-cliques from the Caltech Facebook dataset. Clique edges are shown in red; modularity partitions are shown by color; each node of the 4-clique has been assigned to a different community. Note the many paths of length 2 between the clique’s nodes. This clique will not show up in the list of found structure.

We might be concerned that, as we are only considering the proportion of maximal cliques that have been split, many of the maximal cliques will be small, such as 4-cliques, and that if a 4-clique is split between partitions – while contrary to the traditional intuition of strong ties crossing community boundaries – this might not be of particular concern. For an even more conservative set of experiments, we consider only cliques of at least size 8, with  $\alpha = 0.8$ . These parameters are arbitrary and we do not seek to justify them, other than to reiterate that we are considering conservative structure, which would traditionally be expected only in the ‘cores’ of found communities, not on their boundaries – structure that a comprehensive CFA should return. Even with this conservative definition, the partitions break significant amounts of structure on many networks – see Table 1.

Our results show these proportions of cliques split vary across datasets. There is also a large variation in the number of maximal cliques present. We might reason that this is due to some fundamental difference in the nature of the networks being considered, and question whether such an analysis can be meaningfully applied across a range of networks. After all, the Amazon network is a network of frequently co-purchased products, and the web datasets are explicitly constructed lists of hyperlinks; still other networks involve human communication or collaboration. These networks are, however, frequently treated together as *complex networks*; we might, a priori, expect the same CFAs to perform well on them, and assume that a good CFA in one domain is automatically suitable for another. However, as these results show, this modularity method seems to do poorly on at least some types of network – at least if finding complete lists of structure is desired. Similar results also hold on many networks if we consider just the proportion of  $n$ -cliques split, for a specific value of  $n$ , or if we enforce a Jaccard coefficient distance between cliques we analyse.

### 2.3 Relationship of modularity of partitions to proportion of cliques split

To investigate if the proportion of split cliques is in some way an artefact of low inherent modularity within the networks, we create a scatter-plot of the modularity achieved, against the proportion of maximal cliques split. From Figure 3(a) no obvious relationship appears. Several of the network partitions have high modularity and still display significant clique splitting; if there is a fundamental characteristic that renders particular networks unsuitable for modularity based partitioning, modularity does not capture it.

Table 1: Proportion of maximal cliques split by the Blondel et al. CFA, per network. We show the proportion of maximal cliques, of size 4 or greater, that have more than 10% of their nodes assigned to different partitions. ‘Large cliques split’ is the proportion of maximal cliques, of size 8 or greater, that that have more than 20% of their nodes assigned to different partitions

Network	Vertices	Partitions	Cliques	Cliques split	Large cliques split
amazon0302	262111	173	117054	0.01	0.00
ca-AstroPh	18771	331	27997	0.60	0.32
ca-CondMat	23133	626	8824	0.42	0.15
ca-HepTh	9875	483	2636	0.23	0.00
cit-HepTh	27769	172	419942	0.30	0.06
email-Enron	36692	1363	205712	0.61	0.47
email-EuAll	265009	15743	93267	0.82	0.67
FB-caltech	769	10	31745	0.68	0.27
FB-princeton	6596	21	1286678	0.44	0.22
FB-georgetown	9414	26	1440853	0.41	0.22
ppi-Collins	1622	212	4310	0.16	0.08
Mobile1	10001	182	1550	0.97	0.00
Mobile2	10001	124	3538	0.90	0.00
Mobile3	10001	86	951	0.88	NA
soc-Epinions1	75879	1607	1596598	0.38	0.11
soc-Slashdot0811	77360	771	441941	0.13	0.01
Twitter1	2001	8	23570	0.99	0.66
Twitter2	2001	4	554489	0.15	0.01
Twitter3	2001	7	130399	0.06	0.00
web-NotreDame	325729	693	130965	0.04	0.00
web-Stanford	281903	1013	774555	0.04	0.01
wiki-Vote	7115	30	436629	0.65	0.37

## 2.4 Partitioning using Normalized Edge Cut

Another method that has previously been used for the purpose of community finding, from a different family of algorithms, is the multilevel kernel  $k$ -means partitioning method implemented in *Graclus*[6], that minimises a normalized mincut objective. We examined this method on the same datasets. Unlike the modularity method, which discovers the number of partitions into which to break a graph, *Graclus* requires this to be specified. We would expect a smaller number of partitions would result in a smaller proportion of the maximal cliques being broken, and this effect is visible. However, even when asked to produce a relatively small number of partitions – relative to the network sizes – mincut partitioning results in large proportions of the cliques greater than size 4 being split on many datasets, as shown in Table 2.

## 3 Fundamental partitionability of datasets

Some datasets have a higher proportion of cliques split by partitions than others. This is largely uncorrelated with the mere number of cliques in the dataset, or the number of cliques per node, or per edge, or a number of other simple graph measures, such as clustering co-efficient. After investigating several popular CFAs, we ask whether any partition exists which would not split cliques. Perhaps there were potential partitions that would confine cliques to the cores of the communities found, but these methods were not finding them? To answer this, we consider the network formed by discarding all edges from the network that are not part of at least 4 or 5 cliques. The connected components in this subgraph are the sets of nodes that cannot be placed into separate partitions without splitting any cliques. We calculate the size of the largest connected component of each network, and present this as the proportion of nodes in the network, in Table 2. On some networks, such as Facebook, Twitter, or collaboration networks, any partitioning scheme that is constrained to not split cliques of size five or greater, has to leave the majority of nodes in a single partition. This is an important structural difference between these datasets, and an important result for certain diffusion models of complex contagion[4] which can only spread over strong ties, as it shows strong ties do connect

Table 2: Proportion of cliques split by *Graclus*[6], and *hMETIS*[10], per network. Values shown for 4, 16 and 64 Partitions, with *ufactor* 50, and 16 Partitions with *ufactor*500. Also shown, connected component proportions for subgraphs of edges in at least 4-Cliques, and edges in at least 5-Cliques.

Network	Graclus 4	16	64	hMETIS 4	16	64	16 <i>uf</i> 500	4-Clique	5-Clique
amazon0302	0.01	0.02	0.04	0.00	0.00	0.00	0.00	0.11	0.00
ca-AstroPh	0.43	0.53	0.77	0.27	0.49	0.65	0.34	0.83	0.71
ca-CondMat	0.28	0.40	0.50	0.17	0.30	0.39	0.30	0.71	0.52
ca-HepTh	0.16	0.28	0.43	0.10	0.19	0.29	0.19	0.42	0.13
cit-HepTh	0.13	0.35	0.55	0.15	0.31	0.47	0.30	0.75	0.62
email-Enron	0.38	0.74	0.92	0.10	0.54	0.67	0.38	0.55	0.39
email-EuAll	0.53	0.86	0.98	0.20	0.58	0.76	0.42	0.04	0.02
FB-caltech	0.62	0.86	1.00	0.56	0.89	0.99	0.57	0.89	0.84
FB-princeton	0.33	0.69	0.89	0.32	0.58	0.89	0.36	0.92	0.89
FB-georgetown	0.30	0.58	0.80	0.32	0.50	0.74	0.40	0.93	0.90
ppi-Collins	0.00	0.16	0.93	0.00	0.79	0.95	0.01	0.59	0.36
Mobile1	0.75	0.88	0.99	0.47	0.81	0.93	0.80	0.17	0.07
Mobile2	0.66	0.93	0.97	0.47	0.77	0.92	0.64	0.20	0.09
Mobile3	0.83	0.93	0.96	0.48	0.82	0.95	0.77	0.06	0.01
soc-Epinions1	0.46	0.88	0.81	0.24	0.51	0.63	-	0.18	0.12
soc-Slashdot0811	0.28	0.49	0.94	0.08	0.13	0.37	0.09	0.10	0.04
Twitter1	0.88	0.99	1.00	0.82	0.97	1.00	0.83	0.78	0.57
Twitter2	0.22	0.99	1.00	0.05	0.88	1.00	0.56	0.70	0.57
Twitter3	0.74	0.98	1.00	0.04	0.65	0.99	0.05	0.74	0.33
web-NotreDame	0.01	0.03	0.11	0.00	0.05	0.18	0.04	0.07	0.03
web-Stanford	0.03	0.04	0.39	0.00	0.09	0.46	0.02	0.49	0.40
wiki-Vote	0.48	0.96	1.00	0.51	0.88	0.99	0.51	0.43	0.35

communities. Further, on some of the larger datasets such as Slashdot dataset, with 77,360 nodes, we find that over 30 per cent of those nodes (ie, 23980 nodes) are in a connected component of the subgraph containing only edges that are in *triangles*. This is further evidence against the strict idea that strong ties cannot cross community boundaries.

### 3.1 Partitions that directly minimise clique splits

Having established the limits of partitions that break no single 4-clique, we consider partitioning to directly optimise the number of cliques preserved, while producing balanced partitions. Partitioning a network while splitting as few cliques as possible is a hypergraph partitioning problem, where nodes in a clique together are connected by a hyperedge. This simple observation enables us to use a balanced mincut hypergraph partitioning algorithm, such as implemented by *hMETIS*[10] to partition the graph, while directly minimising clique splitting. *hMETIS* requires an important parameter to determine partition balance. Too high a value results in trivial partitions, with the vast majority of nodes in a single partition; too low might force *hMETIS* to make more aggressive hyperedge cuts than is reasonable. We initially set this *ufactor* at 50 (meaning the largest partition may have 50% larger weight than the average), to allow some unbalance. We examine cuts into 4, 16, and 64 partitions – generally fewer partitions than the modularity maximisation approach finds on these graphs. We also present results for 16 partitions with *ufactor* 500, allowing very large variation in partition size. The results are shown in Table 2. Partitions directly minimising clique split indeed result in reduced proportions of the cliques split, compared to balanced mincut of *Graclus*. As the number of partitions, and balance between partitions, constrain *hMETIS* more than the modularity maximisation method, the results are not directly comparable. However, as this method is directly minimising clique cut, it should approach a lower bound attainable by any partitioning CFA, for the given number of partitions – and, with generous balance parameters, indeed does better than modularity maximisation. Even so, several datasets – notably the collaboration networks, the Wiki voting data, the Facebook and telecoms data – still results in substantial proportions of unrecov-

ered structures, further demonstrating the fundamental global unpartitionability of some empirical networks.

### 3.2 Partitionability of datasets with ground truth

Ground truth community datasets are hard to find for most empirical networks of interest. With protein-protein interaction networks, ground truths are available, as protein complexes that correspond to dense clusters of edges. CFAs, notably  $k$ -clique percolation[18] which does not constrain itself to a partition, and has investigated the overlap in PPI networks, have had success recovering these ground truth communities. We consider the interaction data found in the Combined-AP/MS network [5] using the protein complexes from the CYC dataset of known complexes<sup>3</sup>, as ground truths, to examine how partitionable the ground truth communities are as quantified by our metric. Discarding ground truth communities smaller than triangles, we are left with 126 communities of heterogeneous size. To quantify the severity of this overlap, we treat partitioning the ground truth communities as a hypergraph partitioning problem, where each complex is a hyperedge. We use *hMETIS* to produce a partition of the hypergraph into 126 parts, knowing the large imbalance *ufactor* to be specified ahead of time, and obtaining a min hyperedge cut of 12; approximately 10% of the ground truth communities are split by *hMETIS* such that at least some of their nodes have been assigned to another partition. This shows that even with full knowledge of the communities that we seek to recover, a partitioning must break some ground truth communities; as such, certain naturally occurring networks are not ideally suited for partitioning.

### 3.3 Random and synthetic models of community

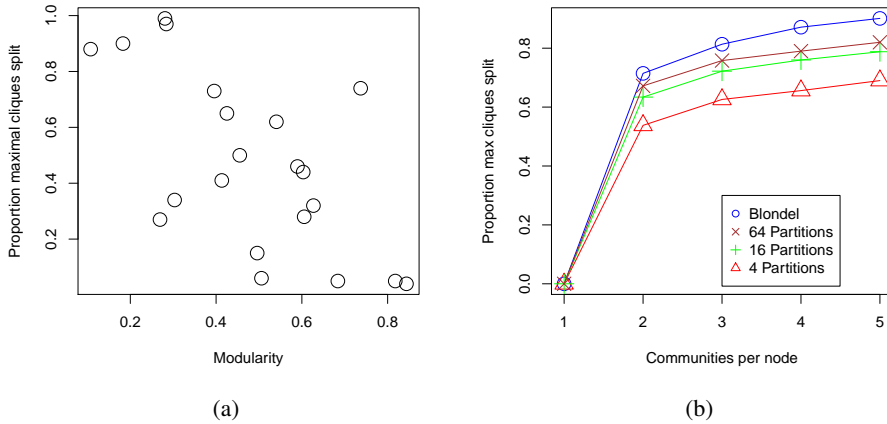


Figure 3: (a) Scatter plot of modularity of the partition vs the proportion of maximal cliques split, for each network. (b) Number of communities-per-node vs proportion of maximal cliques split, by the hypergraph partitioning method, on LFR benchmark data. Each data point is the mean of 5 LFR instances; deviation is negligible.

Broad categories of random community assignment model will produce networks where partitioning will fail to recover full communities. In any random model of community assignment, where the communities are assigned *completely independently*, without an explicit constraint of disjointness, the probability of there existing a partition of the graph that does not split communities is negligible, so long as the average number of communities-per-node in the graph as a whole is greater than 1.0.

Given the following parameters:  $N$ , the total number of nodes in the network;  $s$ , the number of nodes in each individual community; and  $c$ , the number of communities in the network, given a single community, the probability of it being preserved by a random partition is  $1/(2^{s-1})$ . This is conservative, as the probability will be slightly smaller for balanced partitions. For multiple communities, given

<sup>3</sup>Available at <http://wodaklab.org/cyc2008/>

that they be assigned independently of each other, we can see it is  $1/((2^{s-1})^c)$ ; or  $2^{-c(s-1)}$ . Now, there are  $2^{N-1}$  distinct possible partitions, and fewer balanced partitions. Therefore, if  $c \times (s - 1)$  is greater than  $N$ , the chance that there exists a partition respecting all the communities is negligible. This is, to a very close approximation, the same as an average number of communities-per-node of 1.0. Hence, for a random model of community assignment, unless there are explicit constraints that force communities to be disjoint, a partition will almost certainly split some communities.

One source of regarded synthetic benchmark community data is the 'LFR' benchmark[13]. We ran our experiments on 'LFR' graphs to test our method on synthetic data. We generated realisations of a 10,000 node network, altering the number of communities each node was assigned to – from one to five, also increasing the corresponding number of edges, similar to benchmarks detailed in previous work[14]. The results detailing the proportion of cliques split are shown in Figure 3(b). From these results, all methods succeed in partitioning the single community per node networks, but break significant structure on the two or greater community-per-node networks. Even though the synthetic network model isn't directly embedding cliques – just increasing edge density within communities – partitioning fails to find all structure, by our defined metrics, on synthetic networks where nodes are overlapping. Further, large components exist in the graph of edges in cliques in these generated networks. Not only are the individual nodes and communities overlapping as designed by the model; it is a property of the network as a whole that *no* partition exists which does not split cliques.

## 4 Conclusion and further work

We have investigated a wide range of empirical networks, characterising them according to the proportion of cliques that are split by various partitioning methods. Our results show that the early intuition on how communities are embedded in graphs does not hold across all networks and domains. On many complex networks cliques do not exist solely in community cores, connected only by narrow bridges and weak ties – instead they frequently overlap across the community boundaries produced by partitioning algorithms. If we accept cliques as conservative lower bounds for community structure, then, on many networks, partitioning CFAs are fundamentally limited in the completeness of the communities they can find, as shown by our results on graphs of edges in cliques, and hypergraph partitioning. Thus, caution is warranted when using partitioning community finding algorithms where there is a requirement that all significant structures be found. In agreement with recent research on pervasive overlap, conceptualising networks as overlapping meshes of strong ties, but with slightly denser community regions, and using an overlapping CFA may be more appropriate.

This analysis should be extended to individual overlapping community finding algorithms which do not have an in-built partitioning constraint, and examine the lists of structures that they return for completeness. We would like to investigate partitioning the hypergraph of cliques in more detail, exploring different partition parameters. Ideally, we would develop local or sampling based metrics that can attempt to characterise the partitionability of a larger network, as our approach is limited by the need to generate complete lists of cliques – unfeasible on huge networks. Work on formal models of community generation that might explain whether a network is suitable for partitioning, and attempt to characterise the generative processes behind this global overlap would be beneficial. That cliques frequently span communities also has implications for the type of diffusion processes that can occur on networks. We will consider the implications of our results on how we think about diffusion in networks.

## References

- [1] Y.Y. Ahn, J.P. Bagrow, and S. Lehmann. Link communities reveal multi-scale complexity in networks. *arXiv*, 903, 2009.
- [2] 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*, 2008:P10008, 2008.

---

This work is supported by Science Foundation Ireland under grant 08/SRC/I1407: Clique: Graph and Network Analysis Cluster.



- [3] C. Bron and J. Kerbosch. Finding all cliques of an undirected graph. *Communications of the ACM*, 16(9):575–577, 1973.
- [4] D. Centola and M. Macy. Complex contagions and the weakness of long ties. *American Journal of Sociology*, 113(3):702–34, 2007.
- [5] S.R. Collins, P. Kemmeren, X.C. Zhao, J.F. Greenblatt, F. Spencer, F.C.P. Holstege, J.S. Weissman, and N.J. Krogan. Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Molecular & Cellular Proteomics*, 6(3):439, 2007.
- [6] I.S. Dhillon, Y. Guan, and B. Kulis. Weighted graph cuts without eigenvectors a multilevel approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1944–1957, 2007.
- [7] TS Evans and R. Lambiotte. Line graphs, link partitions, and overlapping communities. *Physical Review E*, 80(1):16105, 2009.
- [8] S. Fortunato. Community detection in graphs. *Physics Reports*, 2009.
- [9] M.S. Granovetter. The strength of weak ties. *ajs*, 78(6):1360, 1973.
- [10] G. Karypis and V. Kumar. Multilevel k-way hypergraph partitioning. In *Proceedings of the 36th annual ACM/IEEE Design Automation Conference*, pages 343–348. ACM, 1999.
- [11] B.W. Kernighan and S. Lin. An efficient heuristic procedure for partitioning graphs. *Bell System Technical Journal*, 49(2):291–307, 1970.
- [12] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *Proceedings of the 19th international conference on World wide web*, pages 591–600. ACM, 2010.
- [13] A. Lancichinetti and S. Fortunato. Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Physical Review E*, 80(1):16118, 2009.
- [14] C. Lee, F. Reid, A. McDaid, and N. Hurley. Detecting highly overlapping community structure by greedy clique expansion. *KDD SNA 2010*, 2010.
- [15] J. Leskovec, K.J. Lang, A. Dasgupta, and M.W. Mahoney. Statistical properties of community structure in large social and information networks. In *Proceeding of the 17th international conference on World Wide Web*, pages 695–704. ACM, 2008.
- [16] R.D. Luce and A.D. Perry. A method of matrix analysis of group structure. *Psychometrika*, 14(2):95–116, 1949.
- [17] M.E.J. Newman. Detecting community structure in networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 38(2):321–330, 2004.
- [18] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, 2005.
- [19] A.L. Traud, E.D. Kelsic, P.J. Mucha, and M.A. Porter. Community structure in online collegiate social networks. *arXiv*, 809, 2008.