

RESEARCH

Open Access



# Ensemble clustering for graphs: comparisons and applications

Val  rie Poulin and Fran  ois Th  berge\*

\*Correspondence:

[theberge@ieee.org](mailto:theberge@ieee.org)

Tutte Institute for Mathematics and Computing, Ottawa, Canada

## Abstract

We recently proposed a new ensemble clustering algorithm for graphs (ECG) based on the concept of consensus clustering. In this paper, we provide experimental evidence to the claim that ECG alleviates the well-known resolution limit issue, and that it leads to better stability of the partitions. We propose a community strength index based on ECG results to help quantify the presence of community structure in a graph. We perform a wide range of experiments both over synthetic and real graphs, showing the usefulness of ECG over a variety of problems. In particular, we consider measures based on node partitions as well as topological structure of the communities, and we apply ECG to community-aware anomaly detection. Finally, we show that ECG can be used in a semi-supervised context to zoom in on the sub-graph most closely associated with seed nodes.

**Keywords:** Graph, Clustering, Ensemble, Consensus

## Introduction

Most networks that arise in nature exhibit complex structure (Girvan and Newman 2002; Newman 2003) with subsets of nodes densely interconnected relative to the rest of the network, which we call communities or clusters. Binary relational data-sets are typically represented as graphs  $G = (V, E)$ , where nodes (or vertices)  $v \in V$  represent the entities, and edges  $e \in E$  represent the relations between pairs of entities. Graph clustering aims at finding a partition of the nodes  $V = C_1 \cup \dots \cup C_l$  into good clusters. This is an ill-posed problem (Fortunato and Hric 2016), as there is no universal definition of good clusters, leading to a wide variety of graph clustering algorithms (Girvan and Newman 2002; Clauset et al. 2004; Pons and Latapy 2005; Newman 2006; Raghavan et al. 2007; Reichardt and Bornholdt 2006; Rosvall and Bergstrom 2007; Blondel et al. 2008), with different objective functions. In a recent study (Yang et al. 2016), several state-of-the-art algorithms implemented in the *igraph* (Csardi and Nepusz 2006) package were compared over a wide range of artificial networks generated via the LFR benchmark (Lancichinetti et al. 2008) and some cluster comparison measures. We consider node partitions, also known as non-overlapping communities. Other studies propose methods to compare overlapping communities using cluster comparison measures (Xie et al. 2013) or topological features of the clusters (Orman et al. 2012; Jebabli et al. 2018).

We recently introduced a new ensemble clustering algorithm for graphs (ECG), which compared favorably with leading algorithms (Poulin and Th  berge 2019). The ECG

algorithm is based on the concept of co-association consensus clustering. It is similar to other consensus clustering algorithms such as (Seifi et al. 2013) and in particular (Lancichinetti and Fortunato 2012), but differs in two major points: (1) the choice of an algorithm that alleviates the resolution limit issue for the generation step, and (2) the restriction to endpoints of edges for co-occurrences of node pairs, which keeps low computational complexity.

The contributions in the paper are 4-fold: (1) we provide experimental evidence supporting the claim that ECG alleviates the well-known resolution limit issue of modularity-based algorithms, and that it improves stability compared to the popular Louvain algorithm on which it is based; (2) we introduce a community strength index (CSI) measure based on computed ECG edge weights in order to quantify the presence of community structure in networks; (3) we provide strong evidence of the usefulness of ECG via a wide array of experiments over synthetic and real graphs using several different measures including some of the topological measures proposed in (Orman et al. 2012), and (4) we show that ECG can be used in a semi-supervised context via a "dimmer-like" process to zoom in on important sub-graph(s) given some seed node(s). The rest of the paper is organized as follows. We briefly describe the ECG algorithm, the LFR benchmark and the cluster comparison measures used in the "Background knowledge" section. Some of the advantages of ECG are its stability and its ability to alleviate the well known resolution limit issue. We illustrate those properties in "Resolution limit and stability" section. In the "Weight distribution and community structure" section, we propose a community strength index (CSI) to quantify the presence of community structure in a graph. In the "Experiments" section, ECG is compared to other state-of-the-art algorithms over a wide array of tests, including LFR benchmark and real graphs. We also look at ECG's performance over some measures based on the topological structure of communities. In "Anomaly detection on graphs" section, we re-visit a recently proposed framework (Helling et al. 2019) aimed at finding *anomalous nodes* in graphs using ECG. In "Semi-supervised learning with ECG" section, we show how ECG weights can be used to zoom-in on significant sub-graphs given some seed nodes. We wrap-up in the "Conclusion" section.

## Background knowledge

Let  $G = (V, E)$  be a graph where  $V = \{1, 2, \dots, n\}$  is the set of nodes, and  $E \subseteq \{(u, v) \mid u, v \in V, u < v\}$  is the set of edges. We consider undirected graphs. Edges can have weights  $w(e) > 0$  for each  $e \in E$ . For un-weighted graphs, we let  $w(e) = 1$  for all  $e \in E$ . The 2-core of a graph  $G$  is its maximal subgraph whose nodes have degree at least 2. Let  $P_i = \{C_i^1, \dots, C_i^{l_i}\}$  be a partition of  $V$  of size  $l_i$ . We refer to the  $C_i^j$  as *clusters* of nodes. We use  $\mathbf{1}_{C_i^j}(v)$  to denote the indicator function for  $v \in C_i^j$ .

## The ECG algorithm

The ECG algorithm is a consensus clustering algorithm for graphs. Its generation step consists of independently obtaining  $k$  randomized level-1 partitions from the multilevel-Louvain (ML) algorithm (Blondel et al. 2008):  $\mathcal{P} = \{P_1, \dots, P_k\}$ . Its integration step is performed by running ML on a re-weighted version of the initial graph  $G = (V, E)$ . The

ECG weights are obtained through co-association. The weight of an edge  $e = (u, v) \in E$  is defined as:

$$W_{\mathcal{P}}(u, v) = \begin{cases} w_* + (1 - w_*) \cdot \left( \frac{\sum_{i=1}^k \alpha_{P_i}(u, v)}{k} \right), & (u, v) \in 2\text{-core of } G \\ w_*, & \text{otherwise} \end{cases} \quad (1)$$

where  $0 < w_* < 1$  is some minimum weight and  $\alpha_{P_i}(u, v) = \sum_{j=1}^{l_i} \mathbf{1}_{C_i^j}(u) \cdot \mathbf{1}_{C_i^j}(v)$  indicates if the nodes  $u$  and  $v$  co-occur in a cluster of  $P_i$  or not. When running the ECG algorithm, the size  $k$  of the ensemble and the minimum edge weight  $w_*$  are the only parameters that need to be supplied. Guidelines for the parameters are given in Poulin and Th  berge (2019), where we also show that the results are not too sensitive with respect to those parameters.

### Previous study and the LFR benchmark

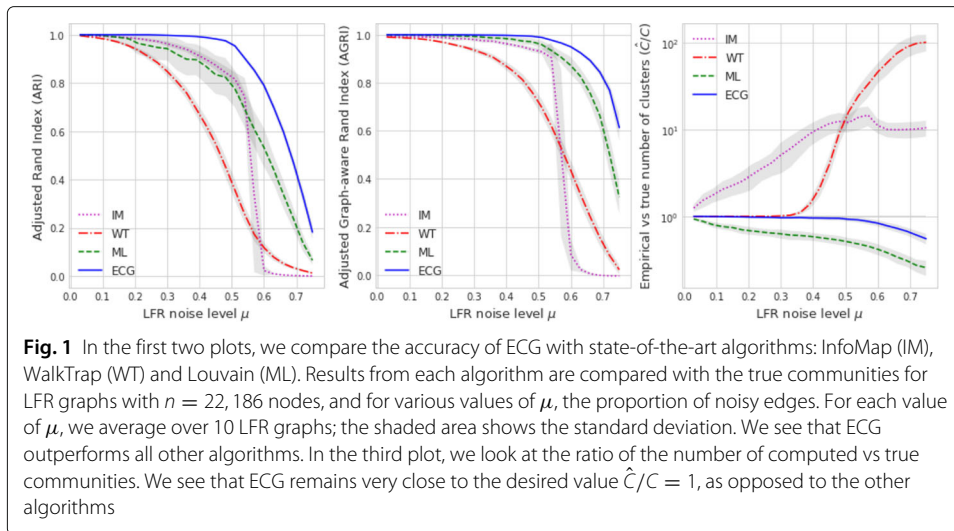
In Poulin and Th  berge (2019), we re-visited a recently published study of graph clustering algorithms, comparing the best performing algorithms from that study with the ECG algorithm. In general, we found ECG to yield better clusters with respect to all of the measures considered. Moreover, ECG generally found a number of communities much closer to the true value. The algorithms are compared on graphs generated with the LFR benchmark for undirected and unweighted graphs and with non-overlapping communities. In the LFR benchmark, three important parameters are: the mixing parameter ( $\mu$ ) which sets the expected proportion of edges for which the two endpoints are in different communities, the (negative) degree distribution power law exponent ( $\gamma_1$ ), and the (negative) community size distribution power law exponent ( $\gamma_2$ ). It is generally recommended to use  $2 \leq \gamma_1 \leq 3$  and  $1 \leq \gamma_2 \leq 2$  to model realistic networks (Lancichinetti and Fortunato 2009; Barabasi 2016). In our previous study, the power law exponents were fixed at  $\gamma_1 = 2$  and  $\gamma_2 = 1$ , with  $.03 \leq \mu \leq .75$ .

### Algorithms and measures

It was shown (Poulin and Th  berge 2018) that graph-agnostic measures such as the adjusted RAND index (ARI) and adjusted mutual information (AMI) (Vinh et al. 2009) yield high scores for *refinements* of the true partition, while a graph-aware version (AGRI) gives high scores for *coarsenings* of the true partition when measuring graph partition similarities. We use both types of measures to compare algorithms. We compared the true communities with those found by the ECG algorithm as well as three other state-of-the-art algorithms: InfoMap (IM) (Rosvall and Bergstrom 2007), WalkTrap (WT) (Pons and Latapy 2005) and multilevel-Louvain (ML) (Blondel et al. 2008). The quality of the results from ECG are clear from the first two plots of Fig. 1, and the number of communities found with ECG remains much closer to the true number as the proportion of noise increases, as shown in the third plot. Those conclusions are illustrative of the results we reported in Poulin and Th  berge (2019).

### Resolution limit and stability

At the heart of ECG is the fact that we use multiple runs of the *single-level* Louvain algorithm to build an ensemble of *weak* (or local) partitionings of the nodes. In this section, we illustrate the two main reasons for this choice.

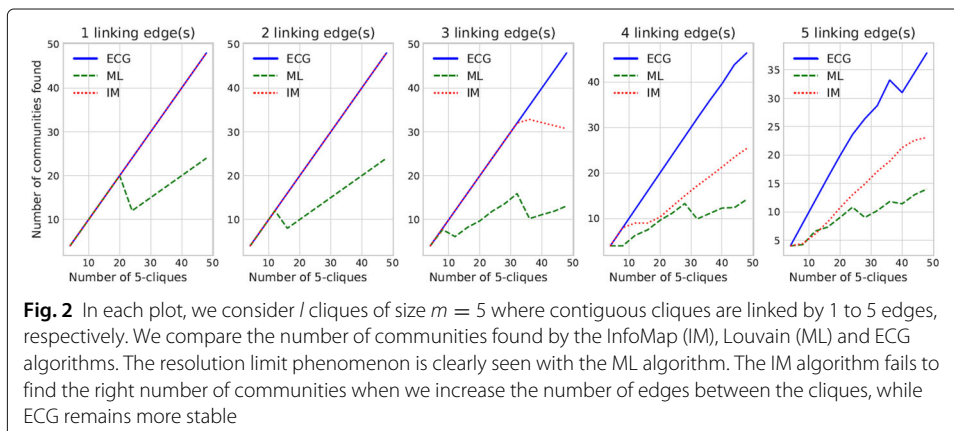


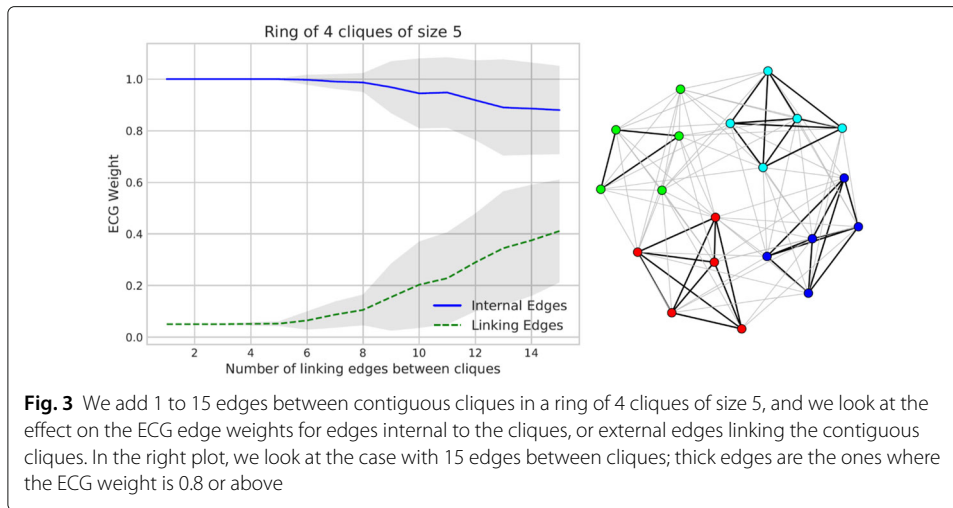
**Resolution issue: ring of cliques illustration**

The resolution limit issue is well illustrated by the infamous ring of cliques example, where the  $n$  nodes form  $l$  cliques (full sub-graphs) of size  $m$ , wired together as a ring. For some choices of  $l$  and  $m$ , grouping pairs of adjacent cliques yields a higher modularity value than the natural choice of each clique forming its own cluster (Fortunato and Barth  lemy 2007). The latter yields higher modularity if and only if  $m(m - 1) > l - 2$ . In (Poulin and Th  berge 2019), we show that choosing a small value for  $w_*$  in (1) can alleviate this issue. In particular, choosing  $w_* < 1/n$  avoids the issue altogether.

In Fig. 2, we look at rings of  $l$  cliques of size  $m = 5$ , with 1 to 5 edges between contiguous cliques. For the ML algorithm, we see the resolution limit issue when  $l > 20$  (with 1 edge between contiguous cliques), which agrees with the known results. The IM algorithm is stable when only a few edges link the cliques, but quickly becomes unstable as more edges are added, while the ECG algorithm remains very stable keeping the default choice of  $w_* = .05$ .

We further illustrate this stability in Fig. 3, where we add up to 15 edges between the cliques of size 5 in a ring with 4 cliques. We see that even when the number of edges linking the cliques is comparable to the number of edges within each clique, the signal





obtained with the ECG weights still favours the cliques. This behaviour allow to better identify communities in noisy graphs. In the right plot of Fig. 3, we show the case where 15 edges are added between contiguous cliques. Thicker edges are the ones where the ECG weights are above 0.8. We see that most of the clique structure is still captured when looking only at those high weight edges.

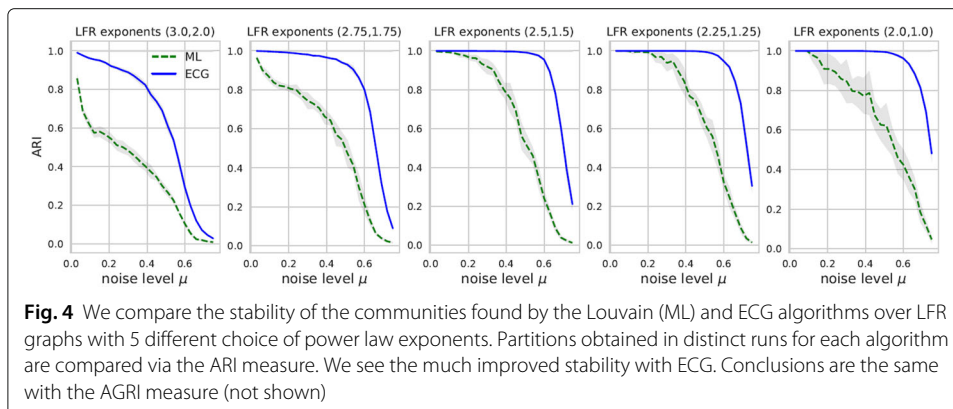
**Stability of ECG**

We illustrate another advantage of ECG which is to significantly reduce the instability in the ML algorithm. To test for stability, we run the same algorithm twice on each graph considered, and we compare the two partitions obtained with the ARI (or AGRI) measure.

In Fig. 4, we did this for the ML and ECG algorithms over LFR graphs with the same parameters as in the previous section. We see that in all cases, ECG greatly improves the stability of the Louvain algorithm.

**Weight distribution and community structure**

When we compare the ECG weight distribution over LFR graphs with varying mixing parameter as well as random graphs, we see that a bi-modal distribution of the weights near the boundaries (0 and 1) is indicative of strong community structure. We



thus propose a simple community strength indicator (CSI) based on the point-mass Wasserstein distance. For all edges  $(u, v) \in E$ , with  $W_{\mathcal{P}}(u, v)$  from (1), we define:

$$CSI = 1 - 2 \cdot \frac{1}{|E|} \sum_{(u,v) \in E} \min(W_{\mathcal{P}}(u, v), 1 - W_{\mathcal{P}}(u, v)) \tag{2}$$

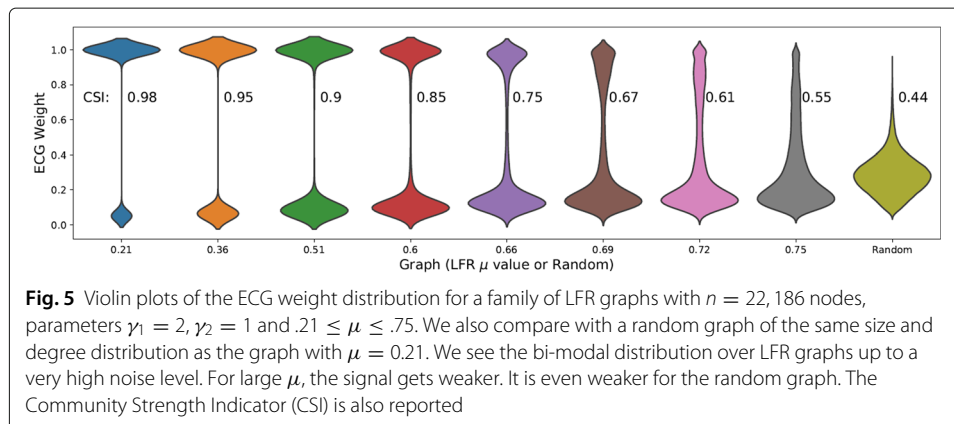
such that  $0 \leq CSI \leq 1$ , where a value close to 1 is indicative of strong community structure, random weights  $W_{\mathcal{P}}(u, v)$  yield a value close to 0.5, and  $CSI = 0$  when all  $W_{\mathcal{P}}(u, v) = 0.5$ . In Fig. 5, we see the bi-modal distribution of the weights for low and mid-range choices of  $\mu$ , along with high CSI values. For larger values of  $\mu$ , the distribution is not as clear, and there are less and less edges with weight close to 1, which indicates a weak community structure, as confirmed by the CSI values. The random graphs have low weights only, which is indicative of the absence of community structure. This example illustrates how the distribution of edge weights obtained with ECG, along with the proposed CSI, can be used to assess the strength of community structure in a graph.

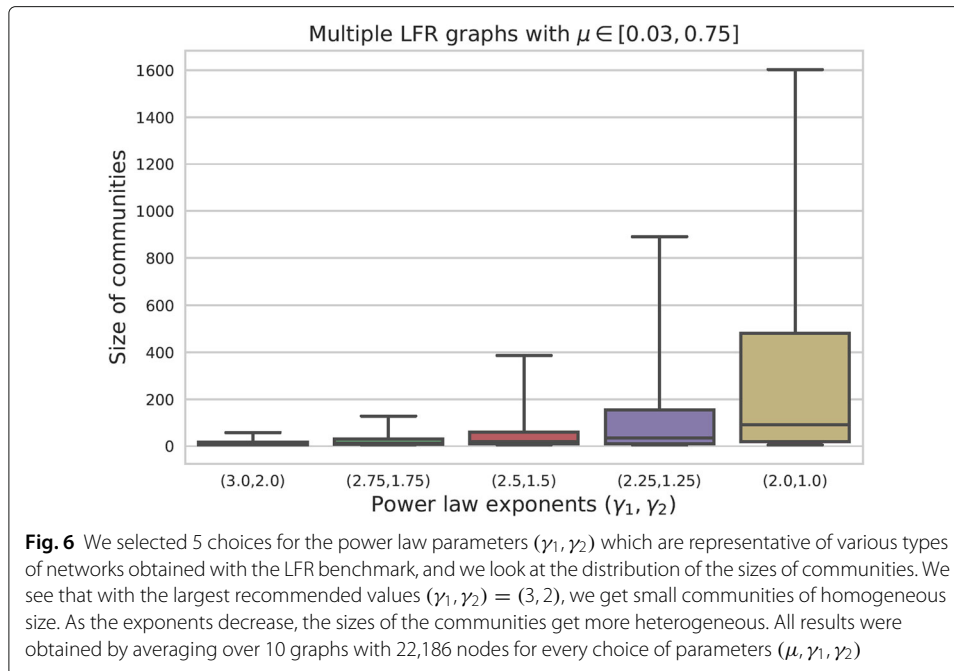
### Experiments

In this section, we experimentally compare ECG to other graph clustering algorithms. First, we consider artificial graphs generated with the LFR benchmark over a choice of parameters which, as we show, yield different community structures. Next, we compare graph clustering algorithms over two real networks with known community structure: a college football graph and a Youtube friendship graph. We further validate the results of ECG by considering some measures based on the topological properties of the communities.

#### Results on LFR benchmark graphs

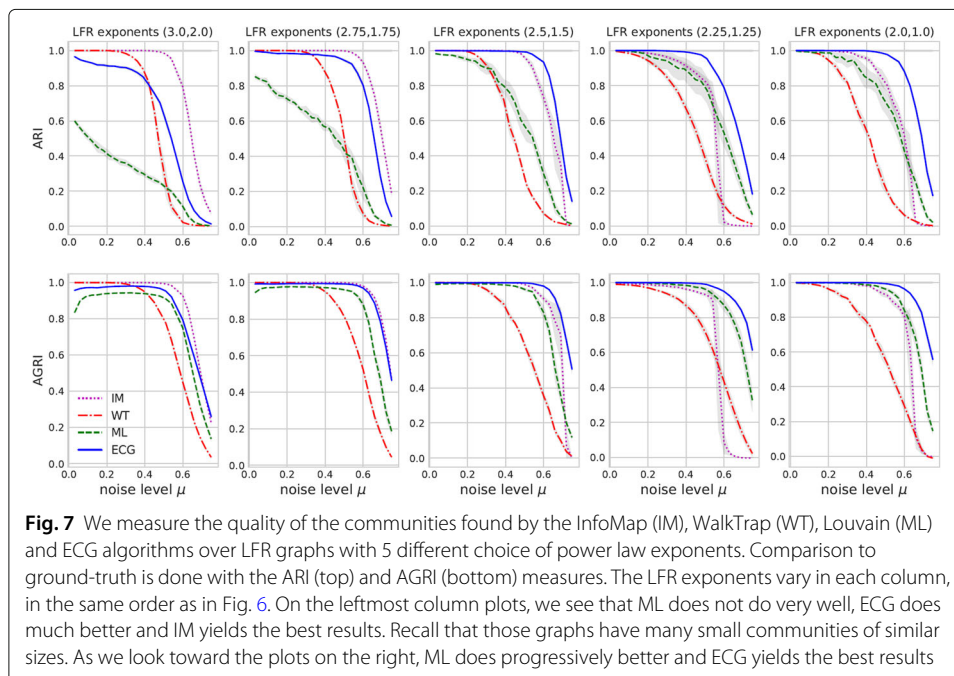
In studies involving LFR benchmark graphs, the power law exponents described earlier are often fixed, while the mixing parameter  $\mu$  is varied to generate graphs with different community strength. However, the choice of power law exponents has strong influence on the type of communities we obtain. In Fig. 6, we show some topological graph differences over 5 choices of parameters  $(\gamma_1, \gamma_2)$  in the recommended range (see Barabasi (2016)). We see that for larger values of those parameters, the communities generated are small and of similar size while smaller values yield graphs with more heterogeneous community sizes.





Thus, considering different values for those parameters amounts to looking at a wider variety of community structures.

In Fig. 7, we compare ECG with IM, WT and ML over a wide range of LFR parameters  $\gamma_1, \gamma_2$  and  $\mu$ , using both the ARI measure and its graph-aware counterpart AGRI. For the larger values of  $(\gamma_1, \gamma_2)$  in the left column, we see that the ML algorithm does not do very well, with ECG doing much better and IM yielding the best results. As the exponents decrease moving toward the right column, we consider graphs with more heterogeneous community size distribution. For those graphs, we see that the ML algorithm



does better, and ECG gives the best results overall. One issue with small communities is that the resolution limit inherent to modularity-based algorithms is more severe. For a ring of size  $m$  cliques, we saw that merging some cliques increases the modularity when  $m$  is small, a special case of “small communities”. The IM algorithm on the other hand is not modularity-based (it uses random walks) and does well in such cases.

### Results on two real networks

We now depart from artificial graphs to look at two real world examples. First, we consider the college football graph studied in Girvan and Newman (2002), which consists of 613 games played between 115 teams which are grouped in 12 conferences (the communities). As noted in Lu et al. (2018), teams generally play more games against other teams in their conference, but there are a few exceptions to this rule. One of the conference is actually a group of independent teams that mainly play against other conferences, and another conference is divided in two groups where most games are within the respective groups. There are also a few outlying teams playing most games with other conference teams. This graph exhibits strong community structure, with average vertex transitivity of .40 and community strength index  $CSI = 0.91$ . The results are summarized in Table 1, where we report the mean results over 100 runs for each algorithm considered earlier. We also report the standard deviation if it is significant (to the third digit). From those results, we see that IM and ECG yield the best results. Moreover, we see that the variance is greatly reduced by using ECG instead of ML, an illustration of the improved stability of ECG already discussed.

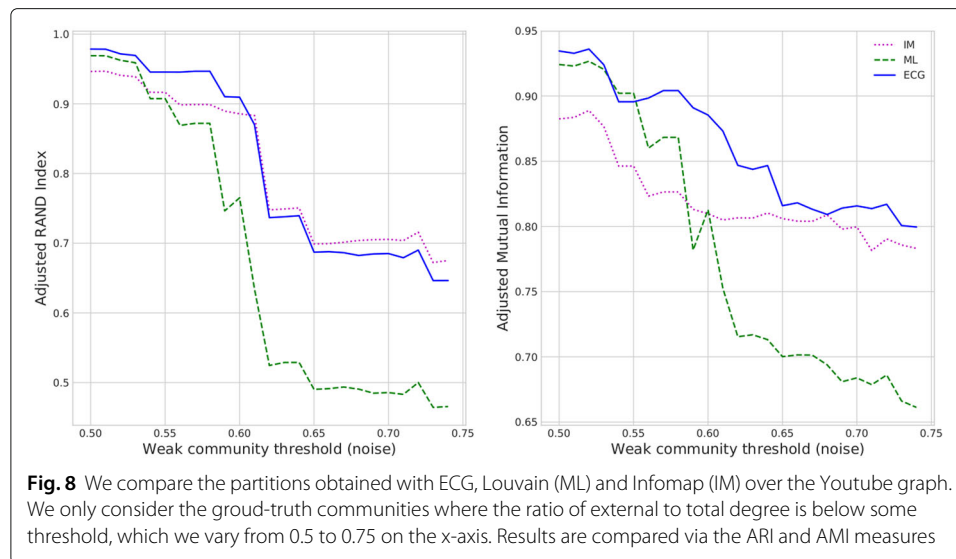
For the next example, we look at the Youtube friendship graph available at (Leskovec and Krevl 2014). There are 1,134,890 nodes (the users) and 2,987,624 edges which consist of friendship between two users. There are 8,385 communities, which are the user-defined groups. The 2-core of this graph spans only about 41.4% of the nodes but nevertheless, it exhibits some community structure with average vertex transitivity of 0.22 in the 2-core, and  $CSI = 0.86$ . The communities (user groups) are however very weak from a topological point of view according to the definition of *weak community* in equation (9.2) of (Barabasi 2016). From this definition, a community  $C$  is a weak community if the ratio of its external degree (edges out of  $C$ ) to its total degree is smaller than 0.5. In the Youtube graph, only 12 communities fulfill this condition. In order to compare algorithms over reasonably coherent communities, we relax the above definition to communities where this ratio is smaller than some *weak community threshold*  $\tau$  for a range of values  $.5 \leq \tau \leq .75$ . This quantity  $\tau$  plays a similar role to the mixing parameter  $\mu$  in LFR benchmark. We apply each clustering algorithm to the entire Youtube graph, except for WT which has complexity  $O(n^2 \log n)$  where  $n$  is the number of nodes. In Fig. 8, we compare the results

**Table 1** We run each clustering algorithm 100 times on the college football dataset, namely: ECG, Louvain (ML), WalkTrap (WT) and InfoMap (IM)

Results for the college football graph				
	ECG	ML	WT	IM
ARI	<b>.889±.016</b>	.763 ± .052	.815	<b>.897</b>
AMI	<b>.900±.005</b>	.843 ± .020	.856	<b>.899</b>
AGRI	<b>.869±.005</b>	.815 ± .019	.837	<b>.872</b>

We compute three different measures with respect to the ground-truth communities (12 conferences). Where significant, we also report the standard deviation. Best results in bold





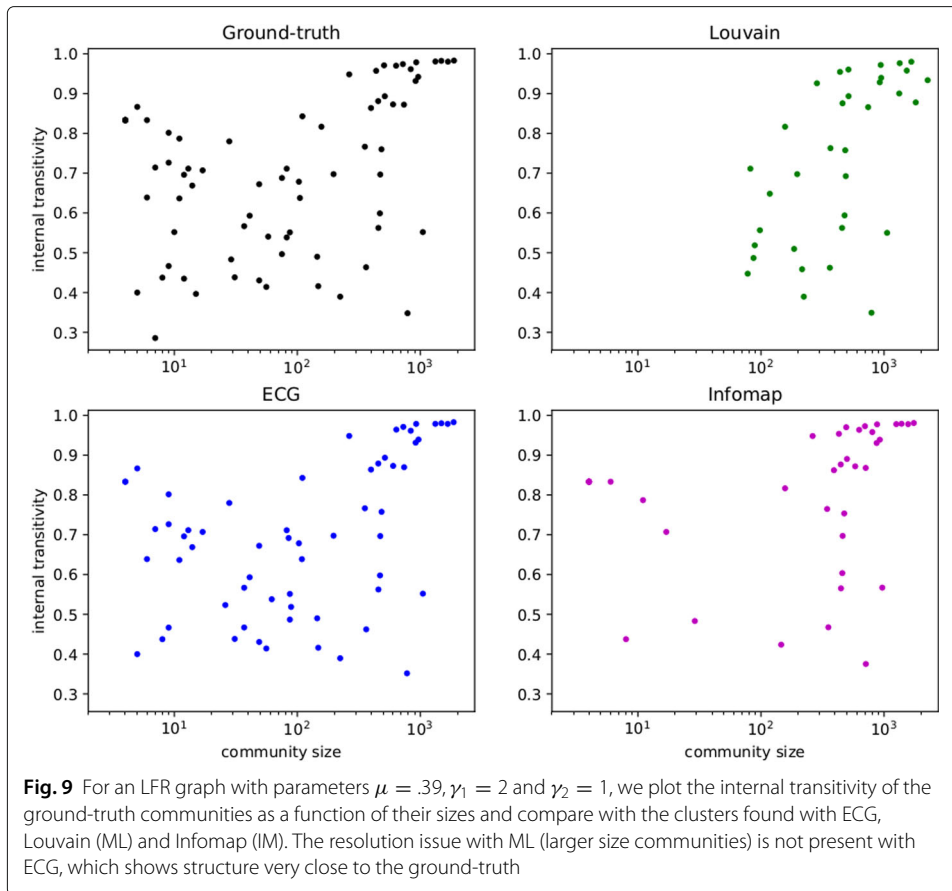
using the ARI and AMI measures. We see that ECG gives very good results in general, with IM also giving good results in particular with respect to the ARI measure. We also see that the performance of ECG decays less rapidly than ML as we saw with LFR graphs, an indication that it is able to capture the local community structure even in the presence of high noise, which was already demonstrated for ring of cliques.

### Topological properties

Quantitative measures such as ARI or AMI are based on the actual clusters found by each algorithm, which are compared to some ground-truth partition. Other types of measures were proposed which are based on topological properties of the clusters; this is useful to ensure that the clusters found by algorithms have structure similar to the real communities. Several such measures are proposed in Orman et al. (2012). We consider two of those measure: the *scaled density* (a variant of edge density), and the *internal transitivity* (based on classic local transitivity). As in Orman et al. (2012), we plot those as a function of the community size for the true communities as well as for the ones found by the different algorithms. Using an LFR graph with parameters  $\mu = .39$ ,  $\gamma_1 = 2$  and  $\gamma_2 = 1$ , we show the results for the internal transitivity measure in Fig. 9. We see that both ECG and to a lesser extent IM follow a distribution similar to the ground-truth communities. For ML, the main issue is that the clusters found are generally larger, an illustration of the resolution limit issue which is much reduced with ECG. In fact, it was shown in Dao et al. (2019) that the distribution of cluster sizes can be a strong indicator of similarity between community detection algorithms. Similar conclusions arise with the scaled density measure, and this plot is available as supplementary material, as well as plots for the college football graph.

### Anomaly detection on graphs

In Helling et al. (2019), the authors propose *CADA*, a community-aware method for detecting anomalous nodes. For each node  $v \in V$ , let  $N(v)$  represent the number of neighbors of  $v$ , and  $N_c(v)$  the number of neighbors of  $v$  that belong to the most represented

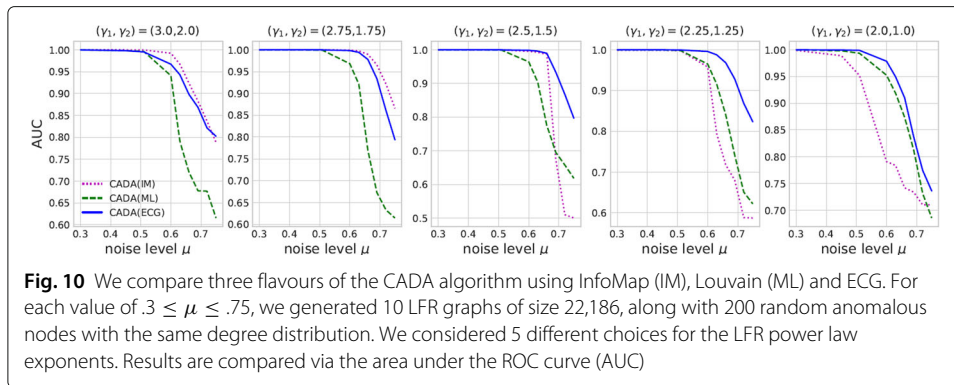


community obtained with the IM or ML algorithm. They define:  $CADA_x(v) = \frac{N(v)}{N_c(v)}$  where  $x \in \{IM, ML\}$  indicates the clustering algorithm used. They compare their algorithm to other methods by generating LFR graphs with degree exponent  $\gamma_1 = 3$  and community size exponent  $\gamma_2 = 2$ . As we saw earlier, this choice corresponds to small communities of homogeneous size, where the ML algorithm performs poorly. We re-visited this approach with ECG, considering different values for the power law exponents. We generated LFR graphs with  $n = 22, 186$  nodes and various values for the mixing parameters. For each graph, we introduced 200 random anomalous nodes with the same degree distribution, as in Fig. 1 of (Helling et al. 2019).

In Fig. 10, we compare  $CADA_{ECG}$  with  $CADA_{IM}$  and  $CADA_{ML}$  using the areas under the ROC curves (AUC). We see that for large choices of the power law exponents, the IM version does best. This is the only choice of parameters used in Helling et al. (2019). As we decrease the values of the exponents, we see that using ECG becomes a better choice, which is supported by our previous results in the “Experiments” section. We also get better results for large values of  $\mu$ , thanks to the increased stability and the ability to distinguish the signal from the noise provided by the ECG weights, which we illustrated earlier in “Resolution limit and stability” section.

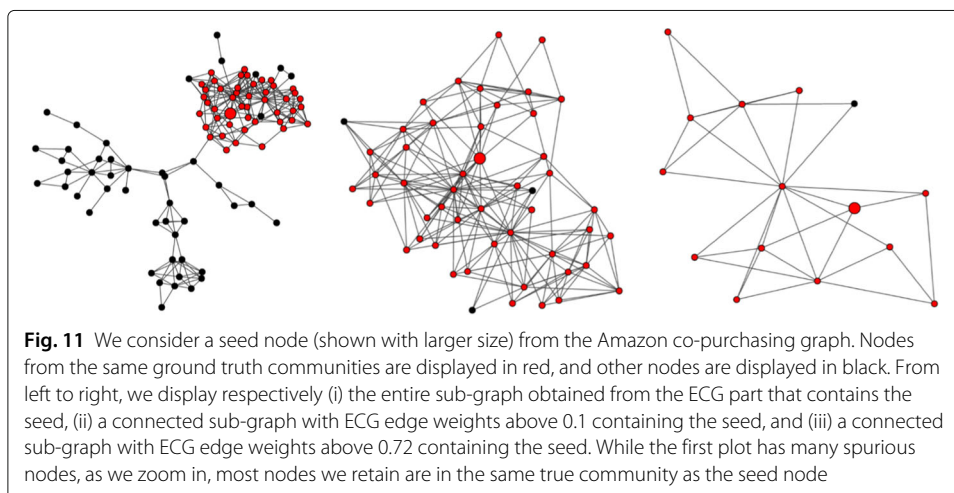
### Semi-supervised learning with ECG

Given some *seed* nodes in a graph, we want to look at the main interactions around those nodes. Taking the seeds’ ego-centered communities is one possibility (Danisch et al.



2013). Another approach is to consider the entire cluster(s) from a partition which contain the seed nodes, but those could be very large. The weights provided by ECG can be used to define a dimmer-like process around the seed nodes, similar to the concept of  $\alpha$ -cores in Seifi et al. (2013), enabling us to highlight the sub-graphs that are the most tightly connected to the seeds. Consider a graph  $G$ , a seed node  $v$  and  $G_v \subset G$  the sub-graph of  $G$  formed by keeping only the ECG cluster containing node  $v$ . Given some threshold  $\theta$ , we delete all edges in  $G_v$  with ECG weights below  $\theta$ , and we keep the connected component sub-graph containing node  $v$ . Increasing  $\theta$  from 0 to 1 provides a hierarchy of sub-graphs of decreasing size which all contain  $v$ .

As an illustration of this process, we consider the Amazon co-purchasing graph available from the SNAP repository (Leskovec and Krevl 2014). This graph has 334,863 nodes and 925,872 edges. There are over 75,000 communities, 5000 of which are identified as the top ones. We picked a node  $v$  that belongs to one of those top communities<sup>1</sup>. We ran ECG, and isolated the sub-graph  $G_v$  induced by the nodes in the ECG cluster that contains  $v$ . In Fig. 11, we gradually increase the threshold  $\theta$ , keeping only edges in  $G_v$  with ECG weight above that threshold, and showing the connected component containing  $v$ . In the first plot, we set  $\theta = 0$ , thus showing  $G_v$  ( $v$  is shown with larger size). Nodes in red belong to the same ground truth community as  $v$ . While we see a lot of spurious nodes in the first plot, discarding edges with low ECG weights (setting  $\theta = 0.1$ ) yields the second



**Table 2** Parameters used for the LFR benchmark graphs

LFR Benchmark Parameters			
$\gamma_1$	$\gamma_2$	avg. degree	max. degree
2.0	1.0	20	2219
2.25	1.25	18	1600
2.5	1.5	15	1000
2.75	1.75	12	304
3.0	2.0	9	149

We generated graphs with  $n = 22,186$  nodes with those choices of parameters, and with mixing parameter  $.03 \leq \mu \leq .75$

sub-graph, where all ground truth nodes are retained. The last plot shows a more aggressive filtering, where we retain only edges with high ECG weights (setting  $\theta = 0.72$ ). This reveals a tightly connected subset around the seed  $v$ .

## Conclusion

In this paper, we provided empirical evidence for two claimed advantages of ECG: its ability to greatly reduce the resolution limit issue of modularity-based algorithms, and its high stability. We also introduced a new index to quantify the presence of community structure in a graph using the ensemble weights in ECG. We validated the above advantages by comparing ECG with state-of-the-art algorithms over a wide range of experiments, including some real graphs and the use of topological features for comparison. We showed ECG to be the best performing algorithm in most cases. Finally, we proposed a framework using ECG in a semi-supervised fashion to extract relevant sub-graphs around seed nodes. The LFR benchmark was used extensively in our experiments. In Orman et al. (2013), two alternatives to the configuration model used in LFR are proposed, and are shown to be more realistic with respect to some topological properties. Those are based respectively on the Barabasi-Albert and the evolutionary preferential attachment models. As future work, we plan to investigate the performance of ECG with respect to those benchmarks.

## Abbreviations

AGRI: Adjusted graph-aware RAND index; AMI: Adjusted mutual information; ARI: Adjusted RAND index; CADA: Community-aware anomaly detection; CSI: Community strength indicator; ECG: Ensemble clustering for graphs algorithm; IM: InfoMap algorithm; LFR: Lancichinetti, Fortunato and Radicchi benchmark; LP: Label propagation algorithm; ML: Multilevel-Louvain algorithm; WT: WalkTrap algorithm

## Acknowledgements

Not applicable.

## Authors' contributions

All authors contributed equally, read and approved the final manuscript.

## Funding

Not applicable.

## Availability of data and materials

The college football graph can be found at (Newman), and the Amazon and Youtube graphs can be found at (Leskovec and Krevl 2014). Code for ECG is openly available (Th  berge and Poulin 2018). All examples using LFR benchmark graphs can be re-created using the code available at (LFR-Benchmark\_UndirWeightOvp). The parameters used for each test were specified in the paper and are listed in Table 2 for reference.

## Competing interests

The authors declare that they have no competing interests.

Received: 19 March 2019 Accepted: 20 June 2019

Published online: 22 July 2019

## References

- Barabasi AL (2016) *Network Science*. Cambridge University Press, UK
- Blondel V, Guillaume JL, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech Theory Exp* 2008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
- Clauset A, Newman M, Moore C (2004) Finding community structure in very large networks. *Phys Rev E* 70(6):066111
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *Intl J Compl Sys* 1695. <http://igraph.org>. Accessed 21 Dec 2018
- Danisch M, Guillaume J-L, Le Grand B (2013) Unfolding ego-centered community structures with "a similarity approach". *Complex Networks IV* 476:145–153
- Dao VL, Bothorel C, Lenca P (2019) Estimating the Similarity of Community Detection Methods Based on Cluster Size Distribution. In: Aiello L, Cherifi C, Cherifi H, Lambiotte R, Li   P, Rocha L (eds). *Complex Networks and Their Applications VII. COMPLEX NETWORKS 2018*. Studies in Computational Intelligence, vol 812. Springer, Cham
- Fortunato S, Barth  lemy M (2007) Resolution limit in community detection. *Proc Nat Acad Sci* 104(1):36–41
- Fortunato S, Hric D (2016) Community detection in networks: A user guide. *Phys Rep* 659:1–44
- Girvan M, Newman M (2002) Community structure in social and biological networks. *Proc Nat Acad Sci* 99(12):7821–7826
- Helling TJ, Scholtes JC, Takes F (2019) A community-aware approach for identifying node anomalies in complex networks. *Compl Netw Appl* VII 1:244–255
- Jebabli M, Cherifi H, Cherifi C, Hamouda A (2018) Community detection algorithm evaluation with ground-truth data. *Physica A Stat Mech Appl* 492(15):651–706
- Lancichinetti A, Fortunato S (2009) Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities. *Phys Rev E* 80(1):016118
- Lancichinetti A, Fortunato S (2012) Consensus clustering in complex networks. *Nat Sci Rep* 2:336
- Lancichinetti A, Fortunato S, Radicchi F (2008) Benchmark graphs for testing community detection algorithms. *Phys Rev E Stat Nonlinear Soft Matter Phys* 78:046110. <https://doi.org/10.1103/PhysRevE.78.046110>
- Leskovec J, Krevl A (2014) SNAP Datasets: Stanford Large Network Dataset Collection. <http://snap.stanford.edu/data>. Accessed 11 Jan 2019
- LFR-Benchmark\_UndirWeightOvp (2009). [https://github.com/eXascaleInfolab/LFR-Benchmark\\_UndirWeightOvp](https://github.com/eXascaleInfolab/LFR-Benchmark_UndirWeightOvp). Accessed Dec 21 2018
- Lu Z, Wahlstr  m J, Nehorai A (2018) Community detection in complex networks via clique conductance. *Sci Rep* 8. <https://doi.org/10.1038/s41598-018-23932-z>
- Newman M (2003) The structure and function of complex networks. *SIAM Rev* 45:167–256
- Newman M (2006) Finding community structure in networks using the eigenvectors of matrices. *Phys Rev E* 74(3):036104
- Newman M American College Football. <http://www-personal.umich.edu/mejn/netdata/>. Accessed 7 May 2019
- Orman GK, Labatut V, Cherifi H (2012) Comparative evaluation of community detection algorithms: a topological approach. *J Stat Mech*. <https://doi.org/10.1088/1742-5468/2012/08/P08001>
- Orman G, Labatut V, Cherifi H (2013) Towards realistic artificial benchmark for community detection algorithms evaluation. *Int J Web Based Comm* 9:349–370. <https://doi.org/10.1504/IJWBC.2013.054908>
- Pons P, Latapy M (2005) Computing communities in large networks using random walks. *Comp Inf Sci ISCIS* 10:284–293. Springer
- Poulin V, Th  berge F (2018) Comparing graph clusterings: Set partition measures vs. graph-aware measures. *CoRR abs/1806.11494*. <http://arxiv.org/abs/1806.11494>
- Poulin V, Th  berge F (2019) Ensemble clustering for graphs. *Compl Netw Appl* VII 1:231–243
- Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. *Phys Rev E* 76(3):036106
- Reichardt J, Bornholdt S (2006) Statistical mechanics of community detection. *Phys Rev E* 74(1):016110
- Rosvall M, Bergstrom CT (2007) Maps of random walks on complex networks reveal community structure. *PNAS* 105(4):1118–1123
- Seifi M, Junier I, Guillaume J-L, Rouquier J-B, Iskrov S (2013) Stable Community Cores in Complex Networks. *Stud Compl Netw* 424. [https://doi.org/10.1007/978-3-642-30287-9\\_10](https://doi.org/10.1007/978-3-642-30287-9_10)
- Th  berge F, Poulin V (2018) Ensemble Clustering for Graphs. <https://www.codeocean.com/>. <https://doi.org/10.24433/CO.0bdd97d9-5f75-4cf4-a797-73151e5aaef4>
- Vinh NX, Epps J, Bailey J (2009) Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In: *Proc. of the 26th Int. Conf. on Machine Learning*. ACM, New York, pp 1073–80. <https://doi.org/10.1145/1553374.1553511>
- Xie J, Kelley S, Szymanski BK (2013) Overlapping community detection in networks: The state-of-the-art and comparative study. *ACM Comput Surv* 45(4):43–14335. <https://doi.org/10.1145/2501654.2501657>
- Yang Z, Algesheimer R, Tessone CJ (2016) A comparative analysis of community detection algorithms on artificial networks. *Nat Sci Rep* 6:30750

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.