

Faster unfolding of communities: Speeding up the Louvain algorithm

V. A. Traag*

Royal Netherlands Institute of Southeast Asian and Caribbean Studies, Reuvensplaats 2, 2311 BE Leiden, the Netherlands and e-Humanities group, Royal Netherlands Academy of Arts and Sciences, Joan Muyskenweg 25, 1096 CJ Amsterdam, the Netherlands

(Received 4 March 2015; revised manuscript received 31 July 2015; published 3 September 2015)

Many complex networks exhibit a modular structure of densely connected groups of nodes. Usually, such a modular structure is uncovered by the optimization of some quality function. Although flawed, modularity remains one of the most popular quality functions. The Louvain algorithm was originally developed for optimizing modularity, but has been applied to a variety of methods. As such, speeding up the Louvain algorithm enables the analysis of larger graphs in a shorter time for various methods. We here suggest to consider moving nodes to a random neighbor community, instead of the best neighbor community. Although incredibly simple, it reduces the theoretical runtime complexity from $O(m)$ to $O(n \log \langle k \rangle)$ in networks with a clear community structure. In benchmark networks, it speeds up the algorithm roughly 2–3 times, while in some real networks it even reaches 10 times faster runtimes. This improvement is due to two factors: (1) a random neighbor is likely to be in a “good” community and (2) random neighbors are likely to be hubs, helping the convergence. Finally, the performance gain only slightly diminishes the quality, especially for modularity, thus providing a good quality-performance ratio. However, these gains are less pronounced, or even disappear, for some other measures such as significance or surprise.

DOI: [10.1103/PhysRevE.92.032801](https://doi.org/10.1103/PhysRevE.92.032801)

PACS number(s): 89.75.Fb, 89.75.Hc, 89.70.Eg

I. INTRODUCTION

Complex networks have gained attention the past decade [1]. Especially with the rise of social media, social networks of unprecedented size became available, which contributed to the establishment of the computational social sciences [2,3]. But networks are also common in disciplines such as biology [4] and neurology [5]. Many of these networks share various common characteristics. They often have skewed degree distributions [6], show a high clustering, and a low average path length [7]. Nodes often cluster together in dense groups, usually called communities. Nodes in a community often share other characteristics: metabolites show related functions [8] and people have a similar background [9]. Revealing the community structure can thus help to understand the network [10].

Modularity [11] remains one of the most popular measures in community detection, even though it is flawed. There have been many algorithms suggested for optimizing modularity. The original algorithm [11] created a full dendrogram and used modularity to decide on a cutting point. It was quite slow, running in $O(n^2m)$, where n is the number of nodes and m the number of links. Many algorithms were quickly introduced to optimize modularity, such as extremal optimization [12], simulated annealing [13,14], spectral methods [15], greedy methods [16], and many other methods [10]. One of the fastest and most effective algorithms is the Louvain algorithm [17], believed to be running in $O(m)$. It has been shown to perform very well in comparative benchmark tests [18]. The algorithm is largely independent of the objective function to optimize, and as such has been used for different methods [19–24].

We first briefly describe the algorithm, and introduce the terminology. We then describe our simple improvement, which we call the random neighbor Louvain, and argue why

we expect it to function well. We derive estimates of the runtime complexity, and obtain $O(m)$ for the original Louvain algorithm, in line with earlier results, and $O(n \log \langle k \rangle)$ for our improvement, where $\langle k \rangle$ is the average degree. This makes it one of the fastest algorithms for community detection to optimize an objective function. Whereas the original algorithm runs in linear time with respect to the number of edges, the random neighbor algorithm is nearly linear with respect to the number of nodes. Finally, we show on benchmark tests and some real networks that this minor adjustment indeed leads to reductions in running time, without losing much quality. These gains are especially visible for modularity, but less clear for other measures such as significance and surprise.

II. LOUVAIN ALGORITHM

Community detection tries to find a “good” partition for a certain graph. In other words, the input is some graph $G = (V, E)$ with $n = |V|$ nodes and $m = |E|$ edges. Each node has k_i neighbors, which is called the degree, which on average is $\langle k \rangle = \frac{2m}{n}$. The output is some partition $\mathcal{V} = \{V_1, V_2, \dots, V_r\}$, where each $V_c \subseteq V$ is a set of nodes we call a community. We work with nonoverlapping nodes, such that $V_c \cap V_d = \emptyset$ for all $c \neq d$ and all nodes will have to be in a community, so that $\bigcup V_c = V$. Alternatively, we denote by σ_i the community of node i , such that $\sigma_i = c$ if (and only if) $i \in V_c$. Both σ and \mathcal{V} may be used interchangeably to refer to the partition. If the distinction is essential, we will explicitly state this.

The Louvain algorithm is suited for optimizing a single objective function that specifies some quality of a partition. We denote such an objective function with \mathcal{H} , which should be maximized. We use $\mathcal{H}(\sigma)$ and $\mathcal{H}(\mathcal{V})$ to mean the same thing. There are various choices for such an objective function, such as modularity [11], Potts models [13,19,22], significance [25], surprise [26], infomap [21], and many more. We will not specify any of the objective functions here, nor shall we discuss

*traag@kitlv.nl

their (dis)advantages, as we focus on the Louvain algorithm as a general optimization scheme.

Briefly, the Louvain algorithm works as follows. The algorithm initially starts out with a partition where each node is in its own community (i.e., $\sigma_i = i$), which is the initial partition. So, initially, there are as many communities as there are nodes. The algorithm moves around nodes from one community to another, to try to improve $\mathcal{H}(\sigma)$. We denote by $\Delta\mathcal{H}(\sigma_i \mapsto c)$ the difference in moving node i to another community c . In particular, $\Delta\mathcal{H}(\sigma_i \mapsto c) = \mathcal{H}(\sigma') - \mathcal{H}(\sigma)$, where $\sigma'_j = \sigma_j$ for all $j \neq i$ and $\sigma'_i = c$, implying that if $\Delta\mathcal{H}(\sigma_i \mapsto c) > 0$, the objective function \mathcal{H} is improved. At some point, the algorithm can no longer improve \mathcal{H} by moving around individual nodes, at which point it aggregates the graph, and reiterates on the aggregated graph. We repeat this procedure as long as we can improve $\mathcal{H}(\sigma)$. The outline of the algorithm is as follows

```

function LOUVAIN (Graph  $G$ )
   $\sigma_i \leftarrow i$ . ▷ Initial partition
   $\sigma' \leftarrow \text{MOVENODES}(G)$  ▷ Initial move nodes
  while  $\mathcal{H}(\sigma') > \mathcal{H}(\sigma)$  do
     $\sigma \leftarrow \sigma'$ 
     $G \leftarrow \text{AGGREGATE}(G, \sigma)$ 
     $\Sigma \leftarrow \text{MOVENODES}(G)$  ▷ Move nodes
     $\sigma'_i \leftarrow \Sigma_{\sigma'_i}$  for all  $i$  ▷ Correct  $\sigma'$  according to  $\Sigma$ 
  end while
  return  $\sigma'$ 
end function

function MOVENODES (Graph  $G$ )
   $\sigma_i \leftarrow i$  for  $i = 1, \dots, |V(G)|$ . ▷ Initial partition
   $q \leftarrow -\infty$ 
  while  $\mathcal{H}(\sigma) > q$  do
     $q = \mathcal{H}(\sigma)$ 
    for random  $v \in V(G)$  do
       $c \leftarrow \text{SELECTCOMMUNITY}(v)$ 
      if  $\Delta\mathcal{H}(\sigma_v \mapsto c) > 0$  then
         $\sigma_v \leftarrow c$ .
      end if
    end for
  end while
end function

function AGGREGATE (Graph  $G$ , Partition  $\sigma$ )
   $A \leftarrow \text{ADJACENCY}(G)$ 
   $A'_{cd} \leftarrow \sum_{ij} A_{ij} \delta(\sigma_i, c) \delta(\sigma_j, d)$ 
  return  $A'$ 
end function

```

There are two key procedures: MOVENODES and AGGREGATE. The MOVENODES procedure displayed in Algorithm 1 loops over all nodes (in random order), and considers moving them to an alternative community. This procedure relies on SELECTCOMMUNITY to select a (possibly) better community c . Only if the improvement $\Delta\mathcal{H}(\sigma_v \mapsto c) > 0$, we will actually move the node to community c . The AGGREGATE procedure may depend on the exact quality function \mathcal{H} used. In particular, the aggregate graph G' should be constructed according to

σ , such that $\mathcal{H}(G', \sigma') = \mathcal{H}(G, \sigma)$, where $\sigma'_i = i$ is the initial partition. That is, the quality of the initial partition σ' of the aggregated graph G' should be equal to the quality of the partition σ of the original graph G . In Algorithm 1 a version is displayed which is suited for modularity. Other methods may require additional variables to be used when aggregating the graph (e.g. [19]).

The only procedure that remains to be specified is SELECTCOMMUNITY. In the original Louvain algorithm, this procedure commonly considers all possible neighboring communities, and then greedily selects the best community. This can be summarized as follows

```

Select best neighbor community
function SELECTCOMMUNITY (Node  $v$ )
   $\delta \leftarrow -\infty$ .
   $c \leftarrow \sigma_v$ .
   $C \leftarrow \{\sigma_u \mid (uv) \in E(G)\}$  ▷ Neighbor communities.
  for Community  $c' \in C$  do
    if  $\Delta\mathcal{H}(\sigma_v \mapsto c') > \delta$  then
       $\delta \leftarrow \Delta\mathcal{H}(\sigma_v \mapsto c')$ 
       $c \leftarrow c'$ 
    end if
  end for
  return  $c$ 
end function

```

We created a flexible and fast implementation of the Louvain algorithm in C++ for use in python using `igraph`. The implementation of the algorithm itself is quite detached from the objective function to optimize. In particular, all that is required to implement a new objective function is the difference when moving a node $\Delta\mathcal{H}$ and the quality function \mathcal{H} itself (although the latter is not strictly necessary). This implementation is available open source from [27] and [28].

III. IMPROVEMENT

Not surprisingly, the Louvain algorithm generally spends most of its time contemplating alternative communities. While profiling our implementation, we found that it spends roughly 95% of the time calculating the difference $\Delta\mathcal{H}(\sigma_v \mapsto c)$ in Algorithm 2. Much of this time is spent moving around nodes for the first time. With an initial partition where each node is in its own community, almost any neighboring community would be an improvement. Moreover, when the algorithm has progressed a bit, many neighbors likely belong to the same community. We therefore suggest that instead of considering all neighboring communities, we simply select a random neighbor, and consider that community (as stated in Algorithm 2), which we call the random neighbor Louvain, as follows

```

Select random neighbor community
function SELECTCOMMUNITY (Node  $v$ )
  return random  $\sigma \in \{\sigma_u \mid (uv) \in E(G)\}$ .
end function

```

Notice that the selection of a random neighbor makes the greedy Louvain algorithm less greedy and thus more explorative. Indeed, when also accepting moves with some probability depending on the improvement (possibly also accepting degrading moves), the algorithm comes close to resemble simulated annealing [13,14]. However, simulated annealing is rather slow for community detection [18], so we don't explore that direction further, since we are interested in speeding up the algorithm.

There are several advantages to the selection of a random neighbor. First of all, it is likely to choose a relatively good community. In general, a node should be in a community to which relatively many of its neighbors belong as well (although this of course depends on the exact quality function). By selecting a community from among its neighbors, there is a good chance that a relatively good community is picked. In particular, if node i has $k_i(c)$ neighbors in community c , the probability that community c will be considered for moving is $k_i(c)/k_i$. The probability for selecting a community is thus proportional to the number of neighbors in that community. Bad communities (with relatively few neighbors) are less frequently sampled, so that the algorithm focuses more on the promising communities (those with relatively many neighbors).

Moreover, when considering the initial partition of each node in its own community, almost any move would improve the quality function \mathcal{H} . The difference between alternative communities in this early stage is likely to be marginal. Any move that puts two nodes in the same community is probably better than a node in its own community. Such moves quickly reduce the number of communities from roughly n to $n/2$. But instead of considering every neighboring community as in the original Louvain algorithm, which takes roughly $O(\langle k \rangle)$, our random neighbor Louvain algorithm only considers a single random neighbor, which takes constant time $O(1)$. So, for the first few iterations, Louvain runs in $O(n \langle k \rangle) = O(m)$, whereas selecting a random neighbor runs in $O(n)$.

Notice there is a big difference between (1) selecting a random neighbor and then its community and (2) selecting a random community from among the neighboring communities. The first method selects a community proportional to the number of neighbors that are in that community, while the second method selects a community uniformly from the set of neighboring communities. Consider for example a node that is connected to two communities, and has $k_i - 1$ neighbors in the first community and only 1 in the other community. When selecting a community of a random neighbor, the probability the good community is considered is $1 - \frac{1}{k_i}$, while the probability is only $\frac{1}{2}$ when selecting a random community.

Secondly, random selection of a neighbor increases the likelihood of quick convergence. The probability that node i is selected as a random neighbor is roughly $k_i/2m$, resembling preferential attachment [29] in a certain sense. Hubs are thus more likely to be chosen as a candidate community. Since hubs connect many vertices, there is a considerable probability that two nodes consider the same hub. If these two (or more) nodes (and the hub) should in fact belong to the same community, chances are high both nodes and the hub quickly end up in the same community.

As an illustration of this advantage, consider a hubs-and-spokes structure, with one central hub and only neighboring spokes that are connected to each other (and always to the hub). So, any spoke node i is connected to nodes $i - 1$ and $i + 1$ and to the central hub, node n . Consider for simplicity that the nodes are considered in order and that every move will be advantageous. The probability that the first node will move to community n is $p_1 = \frac{1}{3}$. For the second node, he will move to community n if he chooses node n immediately (which happens with probability $\frac{1}{3}$), or if he chooses node 1, and node 1 moved to community n , so that $p_2 = \frac{1}{3} + p_1 \frac{1}{3}$. Similarly, for the other nodes $p_i = \frac{1}{3} + p_{i-1} \frac{1}{3} = \sum_{j=1}^i (\frac{1}{3})^j$, which goes to $\frac{1}{2}$ for $n \rightarrow \infty$. This is higher than when just considering a random neighbor community. In that case, the probability the first node will move to community n is still $\frac{1}{3}$. But for the second node, if node 1 moved to community n , only two communities are left: n and 3. In that case, community n is chosen with probability $\frac{1}{2}$. If node 1 didn't move to community n , then node 2 will move to community n with probability $\frac{1}{3}$. In general, node i moves to community n with probability $p_i = p_{i-1} \frac{1}{2} + (1 - p_{i-1}) \frac{1}{3} = p_{i-1} \frac{1}{6} + \frac{1}{3}$. Working out the recurrence, we obtain that $p_i = \frac{1}{3} \sum_{j=0}^{i-1} (\frac{1}{6})^j$, which tends to $\frac{2}{5}$. Selecting a community of a random neighbor thus works better than selecting a random community from among the neighbors. Selecting a community of a random node is even worse. In that case, the probability is $p_i = \frac{1}{n} (1 + \frac{1}{n})^{i-1}$ which tends to 0 for $n \rightarrow \infty$. In short, selecting the community of a random neighbor is likely to choose a new community that will also be chosen by other nodes.

In summary, selecting a random neighbor should work well because of two reasons. First, it tends to focus on communities that are good. Secondly, it should help in convergence because of higher likelihood of selecting hubs. In particular, the evaluation of SELECTCOMMUNITY in the random neighbor Louvain takes a constant time $O(1)$, whereas evaluating all communities takes about $O(\langle k \rangle)$. However, one essential question is whether SELECTCOMMUNITY will not be too frequently evaluated in the random neighbor Louvain to counter this benefit.

To study this question, let us consider a ring of r cliques of n_c nodes each. The cliques [which are complete subgraphs containing $\binom{n_c}{2}$ links] are connected to another clique only by a single link in a circular fashion (i.e., clique i is connected only to clique $i - 1$ and $i + 1$). Most methods tend to find the cliques (or sets of multiple cliques due to the resolution limit [19,30]). Indeed, it is one of the best possible community structures: we cannot add any more internal edges, nor can we delete any external edges without disconnecting the graph. However, for the runtime complexity, the external edges will play only a marginal role. We may therefore simply assume we will work with r disconnected cliques of size n_c . Although the actual runtime will deviate from this, it should provide a reasonable runtime for relatively "clear" communities, and as such provide a lower bound for more difficult communities.

The core question is thus how quickly both the original and the random neighbor Louvain run on cliques. We will assume the clique should become a single community, which is likely to be the case for most methods. Additionally, we assume

$\Delta\mathcal{H} > 0$ only if a node is moved to a larger community, which is likely to be the case for most methods as nodes in a clique have more links to larger communities. The complexity of the original Louvain implementation is simple to evaluate in this case. The first node will be moved to one of its neighbors, an operation that costs n_c evaluations. The second node has only $n_c - 1$ evaluations to make, since the community of the first node disappeared. If we continue in this fashion, the total number of evaluations t is then $\sum_{i=1}^{n_c} n_c - i + 1 = \frac{n_c(n_c+1)}{2} = O(n_c^2)$. The analysis of the expected runtime of the random neighbor Louvain is more difficult (see the Appendix for more details). However, we can provide a lower bound that serves as a rough estimate. Let us again denote by t the total number of operations before the whole clique is identified as a single community. We divide this in different phases of the algorithm, where each phase i runs from the time where there are $n_c - i + 1$ communities, until there are $n_c - i$ communities. In phase 1 we thus start out with n_c communities, and in the next phase there are only $n_c - 1$ communities. If we denote by t_i the number of operation in phase i , then by linearity $E(t) = E(\sum_i t_i) = \sum_i E(t_i)$. Notice that we will only leave phase i whenever a community of size 1 disappears. The probability that a community of 1 disappears is $\frac{n_c-1}{n_c}$, since it will join any other community (except itself). There are at most i communities of size 1 in phase i , so that the probability a community of size 1 is selected is bounded above by $\frac{n_c-i+1}{n_c}$. In fact, such a state is also relatively likely, as the community size distribution tends to become more skewed than a more uniform distribution due to the preferential attachment on the basis of the community sizes. The number of expected operations in phase i is then bounded below by $\frac{n_c}{n_c-i+1}$, and the expected

operations in total is bounded below by

$$E(t) \geq n_c \sum_{i=1}^{n_c} \frac{1}{n_c - i + 1} \tag{1}$$

$$= n_c \sum_{i=1}^{n_c} \frac{1}{i} = O(n_c \log n_c). \tag{2}$$

However, this lower bound gives in fact a very accurate estimate of the expected running time, as seen in Fig. 1. Whereas the original Louvain algorithm runs in $O(n_c^2)$, the random neighbor version only uses $O(n_c \log n_c)$ to put all nodes of a clique in a single community. We used an explicit simulation of this process to validate our theoretical analysis. Running the actual algorithms on cliques yields similar results (Fig. 1).

To get a rough idea of the overall running time, let us translate these results back to the ring of cliques. In that case, we have r cliques of n_c nodes. The runtime for the original Louvain method is $O(n_c^2)$ for each clique, so that the total runtime is about $O(rn_c^2)$. One factor of n_c^2 comes from running over n_c nodes, while the other factor comes from running over $\langle k \rangle \approx n_c$ neighbors. Since $rn_c = n$, and $n\langle k \rangle = m$, we thus obtain an overall running time of Louvain of about $O(rn_c^2) = O(n\langle k \rangle) = O(m)$, similar to earlier estimates [10,17]. Following the same idea, we obtain an estimate of roughly $O(n \log \langle k \rangle)$ for the runtime of the random neighbor Louvain algorithm. So, whereas the original algorithm runs in roughly linear time with respect to the number of edges, the random neighbor algorithm runs in nearly linear time with respect to the number of nodes. Empirical networks are usually

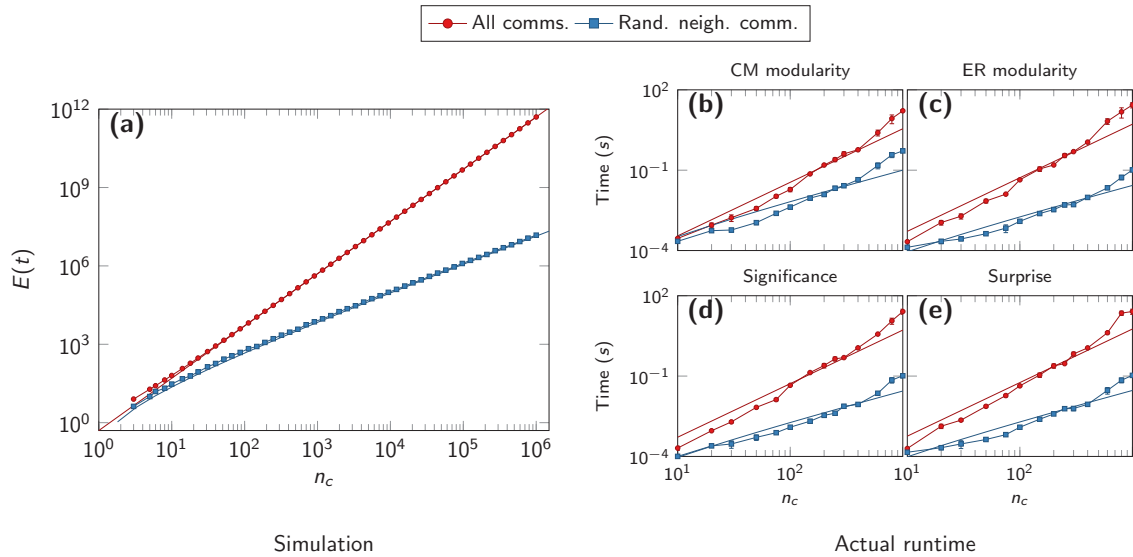


FIG. 1. (Color online) Clique. The original Louvain algorithm considers all communities, which leads to $E(t) = O(n_c^2)$ operations for putting all n_c nodes of a clique in a single community. The improvement considers only random neighbors, which takes only $E(t) = O(n_c \log n_c)$ operations to identify the whole clique as a community. In (a) we show the number of operations in a simulation, with the markers indicating the simulated number of operations, and the solid lines the analytically derived estimates. In (b)–(e) we show the actual time used when optimizing the indicated quality functions for a clique for the different objective functions. The solid lines in (b)–(e) denote best fits to n_c^2 and $n_c \log n_c$ in log space.

rather sparse, so that the difference between $\langle k \rangle$ and $\log(k)$ is usually not that large. Still, it is quite surprising to find such an improvement for such a minor adjustment.

IV. EXPERIMENTAL RESULTS

We use benchmark networks and real networks to show that the random neighbor improvement also reduces the runtime in practice. These benchmark networks contain a planted partition, which we then try to uncover using both the original and the random neighbor Louvain algorithm. An essential role is played by the probability that a link falls outside of the planted community μ . For low μ it is thus quite easy to identify communities, while for high μ it becomes increasingly more difficult. We report results using the speedup ratio calculated as $R_{\text{speed}} = \frac{T_{\text{orig}}}{T_{\text{rn}}}$, where T_{rn} is the runtime of the random neighbor variant and T_{orig} the runtime of the original Louvain method. The runtime is calculated in used CPU time, not elapsed real time. We also report the quality ratio, which is calculated as $R_{\text{qual}} = \frac{\mathcal{H}_{\text{rn}}}{\mathcal{H}_{\text{orig}}}$, where \mathcal{H}_{rn} refers to the quality of the partition uncovered using the random neighbor improvement and \mathcal{H}_{rn} to the quality using the original algorithm. In this way, if $R_{\text{speed}} > 1$ the random neighbor improves upon the original and similarly $R_{\text{qual}} > 1$ if the random neighbor is an improvement. Throughout all plots, error bars indicate standard errors of the mean. The Louvain algorithm can be applied to many different methods, and we here show results for (1) modularity using a configuration null model [11] (CM modularity), (2) modularity using an Erdős-Rényi null model [13] (ER modularity), (3) significance [25], and (4) surprise [31].

We first test the impact of the network size as a whole. We construct benchmark networks ranging from $n = 10^4$ to $n = 10^7$ nodes, with equally sized communities of 1000 nodes, with a Poissonian degree distribution. The speed and quality of the original Louvain algorithm and the random neighbor Louvain algorithm for all four methods is reported in Fig. 2. For all these methods, the random neighbor Louvain speeds up the algorithm roughly 2–3 times. At the same time, the quality of the partitions found remains nearly the same.

However, surprise and significance seem to perform worse than modularity. The speedup is rather limited for higher μ (or becomes even slower than the original), in which case communities are more difficult to detect. Surprise and significance tend to find relatively smaller communities than modularity [26], suggesting that the performance gain of using the random neighbor Louvain is especially pertinent when making a relatively coarse partition. Revisiting the argument of the ring of cliques makes clear that the runtime does not necessarily scale with the degree, but rather, with the clique size, which we may approximate as the community size. Indeed, the runtime for merging all the n_c nodes in a single community together, should take $O(n_c^2)$ originally and $O(n_c \log n_c)$ in the random neighbor Louvain, as previously argued. However, if there are no clear communities present in the network, the running time will not depend on the degree as much, but rather on the sizes of the communities found. Hence the running time should then roughly scale as $O(nn_c)$ for the original implementation and as $O(n \log n_c)$ for the

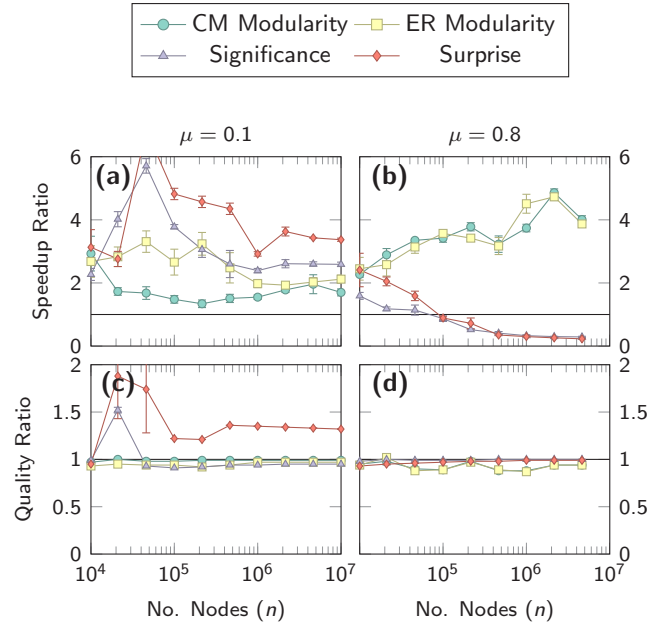


FIG. 2. (Color online) Network size. We here show the ratio of the time and of the quality (i.e., \mathcal{H}) of the uncovered partitions by the original Louvain algorithm and the random neighbor Louvain. The random neighbor Louvain algorithm is 2–3 times faster than the original Louvain algorithm and at some points even faster for clear communities in (a) when using $\mu = 0.1$. However, for less clear communities at $\mu = 0.8$ as displayed in (b), the optimization of significance and surprise is not faster by using the random neighbor Louvain. The random neighbor Louvain uncovers almost the same quality as the original version for a large part, as shown in (c) and (d). However, especially for surprise, the quality is adversely affected by the random neighbor Louvain for $\mu = 0.1$, shown in (c). The results are based on benchmark graph with communities of size $n_c = 1000$ and an average degree of $\langle k \rangle = 15$.

random neighbor Louvain. Since surprise and significance find smaller communities than modularity (unless the communities are clearly defined), the speedup will be rather limited, whereas it will be larger generally for modularity.

We test this by generating benchmark networks with $n = 10^5$ nodes, $\langle k \rangle = 10$ and varying community sizes from 10 to 20 000. Results are displayed in Fig. 3. Indeed, for larger communities, surprise and significance have difficulties discerning such large communities, and it tends to find substructure within these large communities. Notice that modularity also merges smaller communities (thereby uncovering artificially larger communities), part of the problem of the resolution limit [30]. This is exactly also the point at which the speedup for surprise and significance goes down. Moreover, when the community structure is not clear, there is no effect of community size at all. Indeed, in that case, surprise tends to find small communities, and modularity tends to find large communities. The speedup follows this pattern: surprise and significance show very small speedups, while modularity shows larger speedups.

However, modularity also prefers rather balanced communities [32], so that perhaps modularity performs rather well because of the similarity in community sizes. We therefore also consider the impact of more heterogeneity by constructing

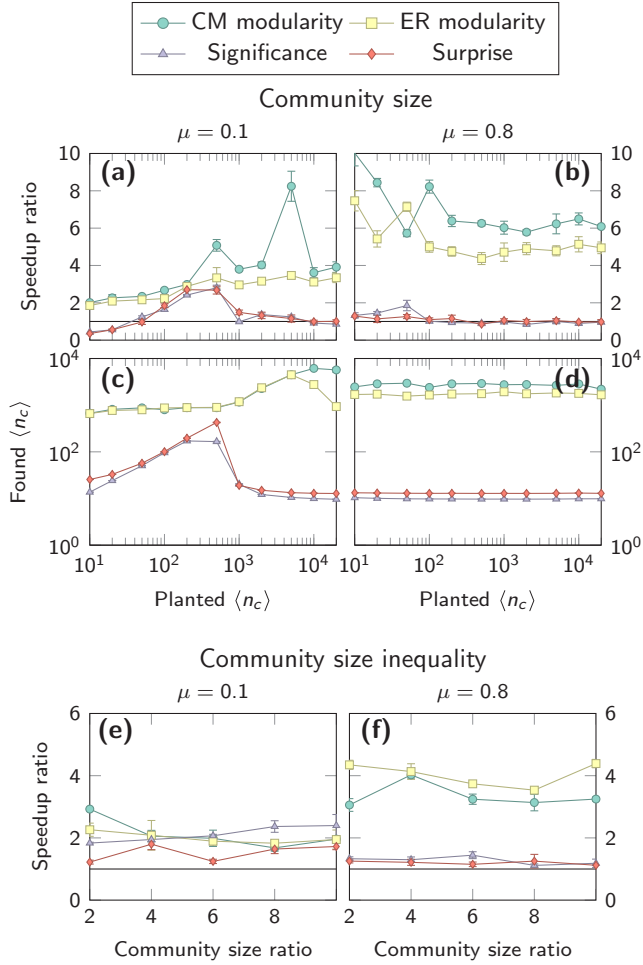


FIG. 3. (Color online) Effect of community size. Results in (a) show that the speedup ratio increases with the community size for $\mu = 0.1$. Surprise and significance find smaller substructures within large communities as seen in (c). This is also when the improvement starts to deteriorate. The results for large communities in (a) and (c) are very similar to the situation when $\mu = 0.8$ in (b) and (d), which resembles a random graph more closely. The speedup ratio in (b) corresponds to this: the speedup is rather large for modularity, while it is much lower for surprise and significance. In (e) and (f) we show that heterogeneity in the community sizes nearly do not impact the speedup ratio. In that case we generate LFR benchmark graphs with smallest community size $n_c = 10$ and the maximum community size varies from 2 to 10 times as large. We use $n = 10^5$ and $\langle k \rangle = 10$ for both benchmarks.

LFR benchmark networks [33]. In these benchmark graphs the community sizes and the degree both follow power law distributions with exponents 1 and 2, respectively. The maximum degree was set at $2.5\langle k \rangle$, while the minimum community size was set at $\langle k \rangle$ for $\langle k \rangle = 10$. We varied the maximum community size from $2\langle k \rangle$ to $10\langle k \rangle$. These results are displayed in Fig. 3, from which we can see that the heterogeneity in community sizes does not affect the results.

We also tested the random neighbor Louvain on six empirical networks of varying sizes. These networks were retrieved from the Koblenz Network Collection [34]. We include (1) the adolescent health data set, a school network collected

TABLE I. Empirical network overview.

Network	n	m	$\langle k \rangle$
Health	2539	12 969	10.22
Brightkite	58 228	214 078	7.35
Facebook	63 731	817 035	25.64
Author Collaboration	22 908	2 673 133	233.38
Web (Google)	875 713	5 105 039	11.66
Web (Berkeley,Stan.)	685 230	7 600 595	22.18

for health research [35], (2) Brightkite, a social network site [36], (3) a Facebook friendship network [37], (4) an author collaboration network from the High Energy topic on arXiv [38], (5) a web hyperlink network released by Google [39], and (6) the complete web hyperlink network from the universities of Berkeley and Stanford [39]. An overview of the size of the networks is provided in Table I, and the results are displayed in Fig. 4. The random neighbor Louvain is clearly faster for most networks and methods, reaching even speedup ratios of over 10 for the hyperlink web network from Berkeley and Stanford. For the web network released by Google the improvement is not faster however. The quality remains relatively similar for most networks, especially for modularity, whereas the quality differs more for surprise and significance.

Notice that significance is not defined for weighted networks, such that significance is not run on those networks (health and author collaboration). But weighted networks raise an interesting point: is it possible to make use of the weight to improve the speed even more? A natural possibility is to sample neighbors proportional to the weight. Neighbors in the same community are often connected with a higher weight, part of the famous strength of weak ties [40,41]. Sampling proportional to the weight should thus increase the chances of drawing a good community. However, this depends on the extent to which this correlation between weight and community holds. The aggregated graph is weighted also, allowing the possibility of weighted sampling as well. On the other hand, only little time is spent in the aggregated iterations, making the benefit relatively small. Weighted sampling in constant time requires preprocessing, which takes an additional $O(m)$ memory and $O(m)$ time. The question is thus whether these costs do not offset the possible benefits.

We use weighted benchmark networks [42] to test whether weighted sampling speeds up the algorithm even further. These benchmark networks introduce an additional mixing parameter for the weight μ_w . Whereas the topological mixing parameter μ controls the probability of an edge outside of the community, the weight is distributed such that on average a proportion of about μ_w lies outside of the community. The strength of the nodes follows the degree $s_i = k_i^\beta$ with $\beta = 1.5$ with $\langle k \rangle = 10$ and $n = 10^5$. The external weight $\mu_w s_i$ is spread over μk_i external links, thereby leading to an average external weight of $\frac{\mu_w s_i}{\mu k_i} = \frac{\mu_w}{\mu} k_i^{\beta-1}$. If $\mu_w > \mu$ the external weight is higher than the internal weight, making it difficult to detect communities correctly. Intuitively, we would thus expect to see an improvement in the random neighbor selection whenever $\mu_w < \mu$, as in that case, the weight correlates with the planted partition. The results for both the unweighted and the weighted

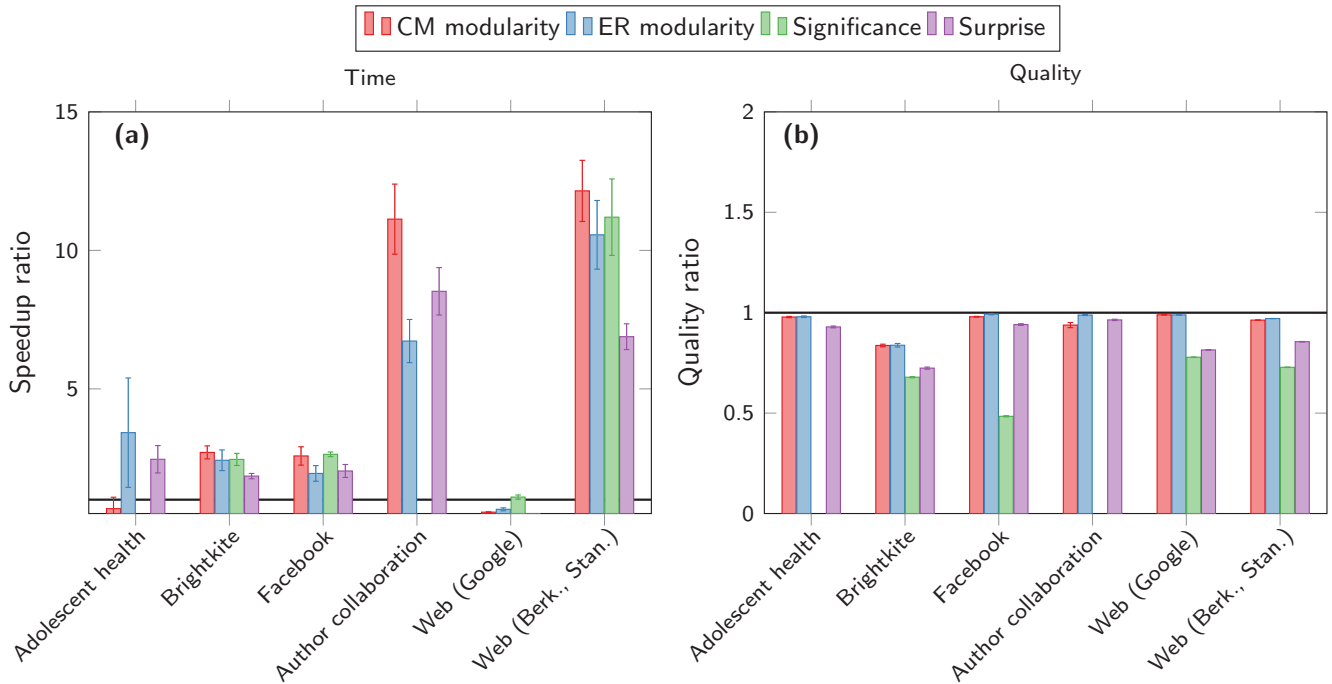


FIG. 4. (Color online) Empirical network results. The random neighbor Louvain usually speeds up the optimization of the objective function for most empirical networks. For the hyperlink network from Google it does not work for any method, while the adolescent health data set poses problems for optimizing CM modularity. The quality remains relatively similar compared to the original, especially for modularity as shown in (b). For significance and surprise the differences are more pronounced.

random neighbor sampling is displayed in Fig. 5. Although the weighted random neighbor sampling sometimes improves on the unweighted variant, overall the performance is comparable. The results on the unweighted benchmark networks and the empirical networks are also very comparable (not shown).

V. CONCLUSION

Many networks seem to contain some community structure. Finding such communities is important across many different disciplines. One of the most used algorithms to optimize some quality function is the Louvain algorithm. We here showed how a remarkably simple adjustment leads to a clear improvement in the runtime complexity. We argue that the approximate runtime of the original Louvain algorithm should be roughly $O(m)$, while the improvement reduces the runtime to $O(n \log(k))$ in a clear community structure. So, whereas the original algorithm is linear in the number of edges, the random neighbor algorithm is nearly linear in the number of nodes.

We have tested the random neighbor algorithm extensively. The improvement is quite consistent across various settings and sizes. The runtime complexity was reduced, speeding up the algorithm roughly 2–3 times, especially when concentrating on the coarser partitions found by modularity. Nonetheless, some methods, such as surprise and significance, are more sensitive to sampling a random neighbor. This seems to be mostly due to the community size in the uncovered partition. Whereas modularity prefers rather coarse partitions, both significance and surprise prefer more refined partitions,

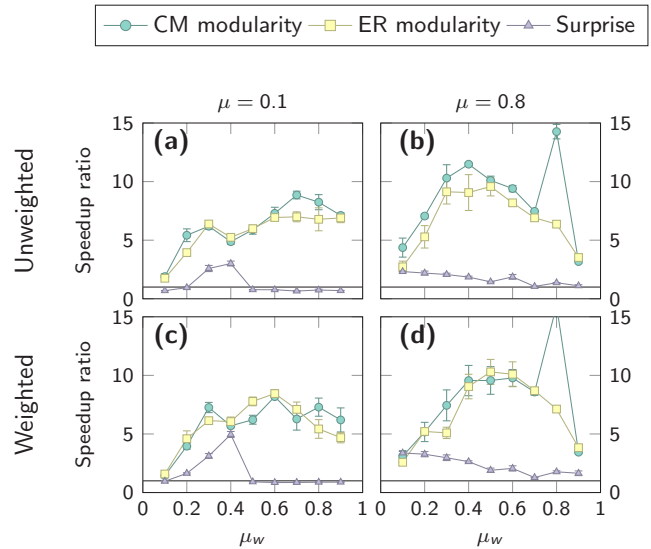


FIG. 5. (Color online) Performance of weighted neighbor sampling. Instead of sampling a neighbor randomly, it is also possible to sample neighbors proportional to the weight. We here test the performance of the unweighted neighbor sampling in (a),(b) and the weighted neighbor sampling in (c),(d). We generate weighted LFR benchmark networks, where the strength of the nodes follows the degree $s_i = k_i^\beta$ with $\beta = 1.5$ with $(k) = 10$ and $n = 10^5$. The results for the unweighted neighbor sampling in (a) and (b) are very similar to the results for the weighted neighbor sampling in (c) and (d). Taking into account the weight hence does not improve the random neighbor sampling much.

leading to much smaller communities. More refined partitions offer fewer opportunities for improving the runtime, so that sampling a random neighbor provides little improvement.

The idea could also be applied in different settings. For example, the label propagation method is also a very fast algorithm [43], but it doesn't consider any objective function. It simply puts a node in the most frequent neighboring community. But instead of considering every neighbor, it can simply choose a random neighbor, similar to the improvement here. We may thus expect a similar improvement in label propagation as for the Louvain algorithm. Similar improvements may be considered in other algorithms. The core of the idea is that a random neighbor is likely to be in a good community, which presumably also holds for other algorithms.

ACKNOWLEDGMENTS

This research is funded by the Royal Netherlands Academy of Arts and Sciences (KNAW) through its eHumanities project [44].

APPENDIX: COMPLEXITY IN A CLIQUE

We here aim to determine the expected number of moves in the random neighbor algorithm. We assume it is always beneficial to move a node to a larger community. In other words, whenever we select a random node i , and a random neighbor j , and the community σ_j of the random neighbor is larger than the community σ_i of node i , i.e., if $|V_{\sigma_j}| \geq |V_{\sigma_i}|$, we will move the node.

Let us denote by f_k the number of communities that have size k ,

$$f_k = |\{c \mid |V_c| = k\}|. \quad (\text{A1})$$

Then $g_k = kf_k$ denotes the number of nodes that belong to a community of size k . Additionally, define $F_k = \sum_{i=k}^n f_i$ the number of communities that have size k or larger. Similarly, define $G_k = \sum_{i=k}^n g_i$ the number of nodes in communities that have size k or larger. Clearly $\sum_k g_k = n$, so that $G_1 = n$. Also, $\sum_k f_k = r$ denotes the number of communities. The probability to select a node from a community of size k is then simply $\frac{g_k}{n}$. Let us denote by X_{cd} the event of moving a node from a community of size c to a community of size d . Then the probability of X_{cd} is

$$\Pr(X_{cd}) = \begin{cases} \frac{g_c}{n} \frac{g_d}{n} & \text{if } c < d, \\ \frac{g_c}{n} \frac{g_c - c}{n} & \text{if } c = d, \\ 0 & \text{if } c > d. \end{cases} \quad (\text{A2})$$

The probability to move it from any community to any other is then $\sum_{cd} \Pr(X_{cd})$. Alternatively, it is easy to see that the probability to move to any other community is $\frac{G_k - k}{n}$ (where we subtract k to make sure it moves to another community, and not the same community). So, overall, the probability we will move a node is then

$$\Pr(\text{move}) = \sum_{k=1}^n \frac{g_k}{n} \frac{G_k - k}{n}. \quad (\text{A3})$$

Similarly, the probability we will not move a node to another community (i.e., we remain stuck in the same partition) is then

$$\Pr(\text{not move}) = \sum_{k=1}^n \frac{g_k}{n} \frac{n - G_k + k}{n}. \quad (\text{A4})$$

We only reduce the number of communities if we move a node from a community of size 1 of course. Hence the probability to reduce the number of communities by 1 is then

$$\Pr\left(\sum_i f_k = r - 1\right) = \frac{g_1}{n} \frac{G_1 - 1}{n}. \quad (\text{A5})$$

Now it would be possible to construct a complete transition network from any of the partitions to other partitions. However, this becomes quickly intractable, and rather difficult to solve.

Instead, we suggest to group partitions by the number of communities. Then, we divide the process into different phases. The algorithm would be in phase i whenever there are $n - i + 1$ communities. In other words, in the first phase there are n communities, and the next phase starts whenever one of these nodes is put in another community. In the penultimate phase, there are only two communities. Let us then denote by t_i the number of moves during phase i , and by t the total number of moves during the whole process. We would like to examine $E(t)$, which we can write out as $\sum_i E(t_i)$ by linearity of expectation.

The number of ways to partition a set of n nodes in r sets can be denoted by $p_r(n)$, which is known as the partition function in number theory [45]. This function obeys the recursive identity

$$p_r(n) = p_r(n - r) + p_{r-1}(n - 1), \quad (\text{A6})$$

since there are $p_{r-1}(n - 1)$ partitions with at least one community of size 1 and $p_r(n - r)$ partitions that all have a community of size at least 2 (since we put r nodes in each one of the r communities). We would then like to know how many partitions there are that have s communities of size 1. Let us first define $q_k(n)$ to denote the number of ways to partition n into k sets with at least one community of size 1. Secondly, we need its counterpart $u_k(n)$ which denotes the number of ways to partition n into k sets without any community of size 1. Obviously then $p_k(n) = q_k(n) + u_k(n)$. We can then derive the recursion

$$q_k(n) = \sum_{r=1}^{k-1} u_{k-r}(n - r) \quad (\text{A7})$$

$$= \sum_{r=1}^{k-1} p_{k-r}(n - r) - q_{k-r}(n - r). \quad (\text{A8})$$

The reasoning is as follows. If there are r communities of size 1, we should know how many partitions there are of $n - r$ nodes into $k - r$ communities without using communities of size 1. Here $q_1(1) = 1$. More specifically, let us denote by $q_k(n, r)$ the number of partitions that have r communities of size 1 and in total k communities, using n nodes. Then, obviously, $q_k(n) = \sum_{r=1}^{k-1} q_k(n, r)$. Moreover, $q_k(n, r) = u_{k-r}(n - r) = p_{k-r}(n - r) - q_{k-r}(n - r)$. The average probability to

reduce the number of communities by 1 is then

$$\Pr\left(\sum f_s = k - 1 \mid \sum f_s = k\right) \quad (\text{A9})$$

$$= \sum_{r=1}^{k-1} \frac{q_k(n,r)}{p_k(n)} \frac{g_1}{n} \frac{G_1 - 1}{n} \quad (\text{A10})$$

$$= \sum_{r=1}^{k-1} \frac{q_k(n,r)}{p_k(n)} \frac{n-r}{n} \frac{n-1}{n}. \quad (\text{A11})$$

This expression is unfortunately not easy to evaluate analytically. Moreover, it incorrectly assumes that each partition is equally likely *a priori*, whereas we know that a more uneven distribution of community sizes is more likely due to the preferential attachment to the largest community. However, the following upper bound is immediate:

$$\Pr\left(\sum f_s = k - 1 \mid \sum f_s = k\right) \leq \frac{n-k+1}{n} \frac{n-1}{n}. \quad (\text{A12})$$

In general, $E(t_i)$ can be calculated relatively straightforward as

$$E(t_i) = \frac{1}{\Pr(\text{reduce community by 1})}, \quad (\text{A13})$$

which using the bound in Eq. (A12) leads to

$$E(t_i) \geq \frac{n}{n-k+1} \frac{n}{n-1}, \quad (\text{A14})$$

so that

$$E(t) = \sum_i E(t_i) \geq \sum_{i=1}^n \frac{n}{n-i+1} \frac{n}{n-1} \quad (\text{A15})$$

$$\approx n \sum_{i=1}^n \frac{1}{i} \quad (\text{A16})$$

$$\approx O(n \log n). \quad (\text{A17})$$

-
- [1] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, *Phys. Rep.* **424**, 175 (2006).
- [2] D. J. Watts, *Nature (London)* **445**, 489 (2007).
- [3] D. Lazer, A. Pentland, L. Adamic, S. Aral, A.-L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, and M. V. Alstynne, *Science* **323**, 721 (2009).
- [4] R. Guimerà, D. B. Stouffer, M. Sales-Pardo, E. A. Leicht, M. E. J. Newman, and L. A. N. Amaral, *Ecology* **91**, 2941 (2010).
- [5] R. F. Betzel, A. Griffa, A. Avena-Koenigsberger, J. Goñi, J.-P. Thiran, P. Hagmann, and O. Sporns, *Netw. Sci.* **1**, 353 (2013).
- [6] A.-L. Barabási, *Science* **325**, 412 (2009).
- [7] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998).
- [8] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási, *Science* **297**, 1551 (2002).
- [9] A. L. Traud, P. J. Mucha, and M. A. Porter, *Physica A* **391**, 4165 (2012).
- [10] S. Fortunato, *Phys. Rep.* **486**, 75 (2010).
- [11] M. E. J. Newman and M. Girvan, *Phys. Rev. E* **69**, 026113 (2004).
- [12] J. Duch and A. Arenas, *Phys. Rev. E* **72**, 027104 (2005).
- [13] J. Reichardt and S. Bornholdt, *Phys. Rev. E* **74**, 016110 (2006).
- [14] R. Guimerà and L. A. Nunes Amaral, *Nature (London)* **433**, 895 (2005).
- [15] M. E. J. Newman, *Phys. Rev. E* **74**, 036104 (2006).
- [16] A. Clauset, M. E. J. Newman, and C. Moore, *Phys. Rev. E* **70**, 066111 (2004).
- [17] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, *J. Stat. Mech.* (2008) P10008.
- [18] A. Lancichinetti and S. Fortunato, *Phys. Rev. E* **80**, 056117 (2009).
- [19] V. A. Traag, P. Van Dooren, and Y. Nesterov, *Phys. Rev. E* **84**, 016114 (2011).
- [20] M. Rosvall and C. T. Bergstrom, *PLoS ONE* **5**, e8694 (2010).
- [21] M. Rosvall and C. T. Bergstrom, *PLoS ONE* **6**, e18209 (2011).
- [22] P. Ronhovde and Z. Nussinov, *Phys. Rev. E* **81**, 046114 (2010).
- [23] T. S. Evans and R. Lambiotte, *Phys. Rev. E* **80**, 016105 (2009).
- [24] A. Lancichinetti, F. Radicchi, J. J. Ramasco, and S. Fortunato, *PLoS ONE* **6**, e18961 (2011).
- [25] V. A. Traag, G. Krings, and P. Van Dooren, *Sci. Rep.* **3**, 2930 (2013).
- [26] V. A. Traag, R. Aldecoa, and J.-C. Delvenne, *Phys. Rev. E* **92**, 022816 (2015).
- [27] <https://github.com/vtraag/louvain-igraph>
- [28] <https://pypi.python.org/pypi/louvain>
- [29] A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [30] S. Fortunato and M. Barthélemy, *Proc. Natl. Acad. Sci. USA* **104**, 36 (2007).
- [31] R. Aldecoa and I. Marín, *Sci. Rep.* **3**, 1060 (2013).
- [32] A. Lancichinetti and S. Fortunato, *Phys. Rev. E* **84**, 066122 (2011).
- [33] A. Lancichinetti, S. Fortunato, and F. Radicchi, *Phys. Rev. E* **78**, 046110 (2008).
- [34] <http://konect.uni-koblenz.de/>
- [35] J. Moody, *Soc. Networks* **23**, 261 (2001).
- [36] E. Cho, S. A. Myers, and J. Leskovec, *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining* (ACM, New York, NY, 2011), pp. 1082–1090.
- [37] B. Viswanath, A. Mislove, M. Cha, and K. P. Gummadi, *Proceedings of the 2nd ACM Workshop on Online Social Networks—WOSN '09* (ACM Press, New York, NY, 2009), p. 37.
- [38] C. F. Jure Leskovec and J. Kleinberg, *ACM Transactions on Knowledge Discovery from Data* (ACM, New York, 2006), pp. 1–40.

- [39] J. Leskovec, K. J. Lang, A. Dasgupta, and M. W. Mahoney, *Proceedings of the 17th International Conference on World Wide Web, WWW'08* (ACM, New York, NY, 2008), pp. 695–704.
- [40] M. Granovetter, *Am. J. Sociol.* **78**, 1360 (1973).
- [41] J. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási, *Proc. Natl. Acad. Sci. USA* **104**, 7332 (2007).
- [42] A. Lancichinetti and S. Fortunato, *Phys. Rev. E* **80**, 016118 (2009).
- [43] U. N. Raghavan, R. Albert, and S. Kumara, *Phys. Rev. E* **76**, 036106 (2007).
- [44] <http://www.ehumanities.nl/computational-humanities/elite-network-shifts/>.
- [45] R. P. Stanley, *Enumerative Combinatorics: Volume 1*, 2nd ed. (Cambridge University Press, Cambridge, UK, 2011).