Modularity-Deltas: a Semi-Global Centrality Measure for Complex Networks

Thomas Magelinski, Mihovil Bartulovic, Kathleen M. Carley
Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA, USA
{tmagelin, mbartulo, carley}@andrew.cmu.edu

Abstract

The phenomena of community structure in real-world networks has been central to network science’s success. While much attention has been paid to community detection algorithms and evaluation criteria, a related question has been overlooked: what would the change in modularity be if a node were to be deleted from a network? In this work, we first show that the answer to this question can be calculated for all nodes in the network in $O(m)$ time. Then, we show that this quantity has implications in three research areas within network science: community analysis, network robustness, and community deception. Modularity-deltas are useful in community analysis as they are a semi-global measure, scalable like local measures but using global information through the network’s partition. In robustness, modularity maximizing attack strategies are more efficient than betweenness-based ones, and are shown to damage the US Powergrid and European Road networks more effectively than degree based attacks. Finally, we show that inverting the robustness method to select modularity minimizing nodes solves the community deception problem in a scalable manner.

1 Introduction

The applications of Network Science span across almost every discipline due to its ability to model complex interacting components in a general way. A key phenomena in Network theory is that of modular networks. That is, networks from a wide array of disciplines exhibit group-like structure. That is, networks often have clusters of nodes that are more tightly connected with each other than to other clusters. Discovery of these clusters have been repeatedly shown shown to be meaningful within their context though empirical studies. Thus, the problem of community detection has been a large part of the field of Network Science.

One way of evaluating a cluster is through modularity. Modularity measures a quality of a partition by comparing the number of links within clusters to that which would fall there by chance. Some of the most successful algorithms rely on maximizing this quantity, namely the Louvain method.

While there has been a great deal of work on algorithms to find communities and methods to evaluate them, a similar question has not been explored: given a graph partition, how would a node’s removal or disappearance effect modularity?

In this work we demonstrate that the question asked above can be answered for all nodes with a calculation that scales as $O(m)$. We further demonstrate that this quantity, which we call modularity-delta, has significant implications for network robustness, community deception, and community analysis itself.
2 Prior Work

2.1 Modularity and Grouping

The most common definition of modularity is that given by Newman, which is the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random \[11\]. We can write modularity \(Q\) of the graph \(G\) as:

\[
Q(G) = \frac{1}{2M} \sum_{v,w} \left( A_{v,w} - \frac{k_v k_w}{2M} \right) \delta(c_v, c_w),
\]

where \(A\) is the adjacency matrix, \(k_v\) is total degree of node \(v\) and \(\delta(a, b)\) is the Kronecker delta function. This modularity is also known as Newman modularity.

The most commonly used community detection algorithms seek to maximize this quantity. Because it is an NP-hard problem, many different methods have been proposed to varying degrees of success \[4, 1, 12\]. The Louvain method has prevailed for years, and has been empirically shown to give meaningful communities a countless times \[1\].

However, recently, Traag, Waltman, van Eck have shown a flaw in the Louvain method \[12\]. Because of its update step, Louvain does not guarantee that its communities are internally connected. It was shown that in fact, many communities are often not connected when using the method on real-world datasets. Instead, they have proposed Leiden grouping, which is slightly faster than Louvain, guarantees well-connected communities, and often achieves higher modularity. As such, we use Leiden grouping.

2.2 Network Centrality Measures

A fundamental question in network science is, which nodes are important? Many metrics and centrality measures have been created to answer this question in different ways \[13\]. Some measures are local, meaning that each node’s value is only dependent on its neighbors. These metrics are highly scalable but have limited value in explaining more complex structure. The most common of these measures is node degree, or the number of neighbors a node has. More powerful measures are global, which incorporate structure beyond just a node’s neighborhood. A common example is betweenness, which is the number of shortest paths that pass through a node. Some of these metrics are very computationally expensive, and are thus intractable for large graphs.

No centrality measures that we are aware of take into account the group structure of a graph. However, this is a common question to ask about a network; given this grouping, which nodes are most important for each group? The best approach to this question currently, is to run a normal centrality measure, and then high centrality nodes from each group.

In this work, we will show that modularity-deltas bridge the gap between local and global centrality measures. They are as easy to calculate as local measures, but can obtain global information about the network through its partitioning. In this sense, it can be thought of as a “group-centrality” measure.

2.3 Network Robustness

Network Robustness refers to how a network responds to attacks. Attacks typically take the form of removal of edges or removal of nodes. We will focus on removal of nodes. Understanding how networks react with missing nodes or edges has important implications in many fields, including but not limited to biology and ecology.

To understand how attacks affect a network, different types of attacks have been developed. In general, a centrality measure is calculated for each of the nodes, and the node with the highest centrality is attacked. Holme introduced this idea along with two styles of attacks: initial and recomputed \[6\].

In the initial case, centralities are calculated once and the top-k nodes are removed. In the recomputed case, centralities are recomputed each time a node is removed. This makes the attack more expensive to compute, but can be more effective.

In this framework, attacks are defined by two characteristics, the centrality measure and the style. Common choices of centrality measure are degree and betweenness. Betweenness has been shown
to be much more damaging to a network, but is far more expensive to compute\cite{6,5}. The shorthand for these methods are based on the acronym of the centrality and style; IB means an attack using initial calculation of betweenness centrality, while RD is recomputed degree.

One method of evaluating an attack’s effectiveness is through network fragmentation. Fragmentation, $\sigma$ can be defined as the size of the remaining largest component $N_\rho$ relative to the initial size of the graph, $N$, where $\rho$ is the fraction of nodes removed. Fragmentation can then be given as $\sigma(\rho) = \frac{N_\rho}{N}$. This is a useful measure because networks often rely on connectivity to function properly. Disconnected components in biological, communication, or power-grid networks are serious danger of failing completely.

A connection between the modular structure in networks and their robustness has been illustrated in \cite{5}. The authors developed a slightly more complex attack strategy which is able to fragment real-world networks far more quickly than the simple methods previously described. They achieve this by insuring that nodes are attacked only when they are in the largest component and when they are connecting groups.

Though effective, betweenness-based attacks do not scale to the size of networks commonly seen on social media. For weighted networks, a single calculation of betweenness scales as $O(nm + n^2 \log n)$, making RB scale as $O(n^2 m + n^3 \log n)$ \cite{2}. This makes RB intractable for medium-sized networks, which is why da Cunha et al. use IB as the base for their attack method \cite{5}. However, even IB is intractable for very large networks.

In this work, we suggest that modularity-deltas bridge the gap between degree and betweenness. It is a semi-global centrality, since it uses global group information, but it is as scalable as local measures such as degree.

### 2.4 Community Deception

Community Deception has recently been formalized by Chen et al \cite{3}. They argue community detection is a very powerful tool, and could potentially be too powerful for privacy-sensitive applications. In order to prevent data sensitive data that is easily identifiable, community structure should be obscured. They then develop a genetic algorithm that attempts to determine which nodes or edges to hide such that the modularity is decreased the most. Genetic algorithms, however, are extremely complex and thus do not scale. Because of this, experiments were only performed on networks with approximately 100 nodes.

Modularity-deltas effectively solve this problem. Given that it is a combinatorial optimization problem, we cannot expect a true solution. However, flipping the robustness problem around gives an accurate and cheap greedy solution. Instead of attacking nodes which will increase modularity the most, nodes are attacked which decrease modularity the most. We demonstrate the power of this approach by performing community deception on true social-media-scale networks.

### 3 Calculating Modularity Deltas

**Definition 3.1.** Let $G = (V,E)$ be a graph where $V$ is a set of vertices, and $E$ of links where $|E| = M$ is total number of links in the graph $G$. Let us define modularity as the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random \cite{11}. We can write modularity $Q$ of the graph $G$ as:

$$Q(G) = \frac{1}{2M} \sum_{v,w} (A_{v,w} - \frac{k_v k_w}{2M}) \delta(c_v, c_w)$$ (2)

$A$ the adjacency matrix, $k_v$ total degree of node $v$ and $\delta(a, b)$ is the Kronecker delta function. This modularity is also known as Newman modularity \cite{11} and we will work with it for the in the following two theorems.

**Theorem 3.2.** Given the modularity defined as in Equation\cite{7}, the difference in modularity between graphs $G$ and $G_\Delta$ can be written as:
\[ \Delta Q(G, G_n) = \frac{m_G - m_n}{M - k_n} - \frac{m_g}{M} - \frac{1}{4M^2} \sum_{c \in \mathcal{C}} \left( h_c^2 - 2h_c \sum_{v \in N\backslash \{n\}} k_v \delta(c_v, c_n) \right) \]  

Where \( G_n \) denotes graph \( G \) with node \( n \) removed, \( \mathcal{C} \) is the set of all communities, \( m_G \) denotes the total number of links within groups, \( m_n \) is the number of links within groups from node \( n \), and \( h_c = m_{n,c} + k_n \delta(c, c_n) \) where \( m_{n,c} \) is the total number of connections from node \( n \) to community \( c \). (Note that \( m_n = m_{n,c} \) if we are looking at community that node \( n \) is part of)

**Proof.** Let us begin by rewriting the definition of modularity so both the actual modularity \( Q_{act}(G) \) and expected modularity \( Q_{exp}(G) \) are explicitly seen:

\[ Q(G) = \frac{1}{2M} \sum_{v,w} \left( A_{v,w} - \frac{k_v k_w}{2m} \right) \delta(c_v, c_w) \]

\[ = \frac{m_G}{M} - \frac{1}{4M^2} \sum_{v,w} k_v k_w \delta(c_v, c_w) \]

After removing node \( n \) the new \( Q_{act}(G_n) \) can be written as:

\[ Q_{act}(G_n) = \frac{m_G - m_n}{M - k_n} \]

To obtain the new \( Q_{exp}(G_n) \) we need to break it down to a sum over all communities. Further, each node multiplies its degree by the sum of the degrees of its community members, \( k_v \sum_{c_v=c_w} w k_w \).

Consider a community, \( c \), such that removed node \( n \) does *not* belong to.

**Case 1.** If you are *not* a neighbor of \( n \):

\[ k_v \sum_{c_v=c_w} w k_w \rightarrow k_v (\sum_{c_v=c_w} w k_w - m_{n,c}) \]

\[ = k_v \sum_{c_v=c_w} w k_w - k_v m_{n,c} \]

where \( m_{n,c} \) is the total number of connections from \( n \) to community \( c \).

**Case 2.** If you *are* a neighbor of \( n \), your degree changes as well:

\[ k_v \sum_{c_v=c_w} w k_w \rightarrow (k_v - 1)(\sum_{c_v=c_w} w k_w - m_{n,c}) \]

\[ = k_v \sum_{c_v=c_w} w k_w - \sum_{c_v=c_w} w k_w - k_v m_{n,c} + m_{n,c} \]

Summing over \( v \), noting that there are \( m_{n,c} \) instances of the second case:

\[ \sum_{v,w \in C} k_v k_w \rightarrow \sum_{v,w \in C} k_v k_w - m_{n,c} \sum_{w} k_w - \sum_{v} k_v m_{n,c} + m_{n,c}^2 \]

\[ = \sum_{v,w \in C} k_v k_w - 2m_{n,c} \sum_{v} k_v + m_{n,c}^2 \]

Next, we have two cases when \( c \) *does* contain \( n \). These cases are different because network loses both the internal connections, and the node’s degree.
Case 3. If you are a not a neighbor of $n$:

$$k_v \sum_{w: c_v = c_w} k_w \rightarrow k_v \left( \sum_{w: c_v = c_w} k_w - m_{n,c} - k_n \right),$$

$$= k_v \sum_{w: c_v = c_w} k_w - k_v m_{n,c} - k_v k_n$$  \hspace{1cm} (9)

Case 4. If you are a neighbor:

$$k_v \sum_{w: c_v = c_w} k_w \rightarrow (k_v - 1) \left( \sum_{w: c_v = c_w} k_w - m_{n,c} - k_n \right),$$

$$= k_v \sum_{w: c_v = c_w} k_w - k_v m_{n,c} - k_v k_n - \sum_{c_w = c_w} k_w + m_{n,c} + k_n$$  \hspace{1cm} (10)

Case 5. And for $n$ itself:

$$k_v \sum_{w: c_v = c_w} k_w \rightarrow 0 \left( \sum_{w: c_v = c_w} k_w - m_{n,c} - k_n \right) = 0,$$

Now summing these together over $v$:

$$\sum_{v,w \in E} k_v k_w \rightarrow \sum_{v,w \in E} k_v k_w - k_n \sum_{c_v = c_w} k_w - m_{n,c} \left( \sum_{v} k_v - k_n \right) - k_n \left( \sum_{v} k_v - k_n \right) - m_{n,c} \sum_{c_w = c_w} k_w + m_{n,c}^2 + m_{n,c} k_n$$

$$= \sum_{v,w \in E} k_v k_w - 2k_n \sum_{c_v = c_w} k_v - 2m_{n,c} \sum_{c_w = c_w} k_v + 2m_{n,c} k_n + k_n^2 + m_{n,c}^2$$  \hspace{1cm} (11)

$$= \sum_{v,w \in E} k_v k_w - 2(m_{n,c} + k_n) \sum_{c_w = c_w} k_v + (m_{n,c} + k_n)^2$$  \hspace{1cm} (12)

If we define $h_c = m_{n,c} + k_n \delta(c, c_n)$, then the change in a community is given by:

$$\sum_{v,w \in c} k_v k_w \rightarrow \sum_{v,w \in c} k_v k_w - 2h_c \sum_{v} k_v + h_c^2$$

Thus the total change in modularity is:

$$\Delta Q(G, G_n) = \frac{m_g - m_n}{M - k_n} - \frac{m_g}{M} - \frac{1}{4M^2} \sum_{c \in \mathcal{E}} \left( h_c^2 - 2h_c \sum_{v \in \mathcal{N} \setminus \{n\}} k_v \delta(c_v, c_n) \right)$$  \hspace{1cm} (13)

**Theorem 3.3.** If we remove node $n$ from the graph $G$ then the new modularity of the new graph $G_n$ can be written as:

$$Q(G_n) = \frac{m_g - m_n}{M - k_n} - \frac{1}{4(M - k_n)} \sum_{c \in \mathcal{E}} (d_c - h_c)^2$$  \hspace{1cm} (14)

where $d_c = \sum_{v \in c} k_v$, $h_c = m_n + k_n \delta(c, c_n)$, and $m_n = \sum_{w} A_{n,w} \delta(c_n, c_w)$
Proof. From the definition of modularity we can write:

\[ Q(G) = \frac{1}{2M} \sum_{v,w} \left[ A_{v,w} - \frac{1}{2M} k_v k_w \right] \delta(c_v, c_w) \]

\[ = \frac{1}{2M} \sum_{c} \sum_{v,w \in c} \left[ A_{v,w} - \frac{1}{2M} k_v k_w \right] \]

\[ = \frac{1}{2M} \sum_{c} \sum_{v \in c} \sum_{w \in c} A_{v,w} - \frac{1}{4M^2} \sum_{c} \sum_{v \in c} \sum_{w \in c} k_v k_w \]

Where \( m_c \) denotes total number of edges that fall within the given groups.

\[ Q(G) = \frac{m_c}{M} - \frac{1}{4M^2} \sum_{c} \sum_{v \in c} \sum_{w \in c} k_v k_w \quad (19) \]

\[ = \frac{m_c}{M} - \frac{1}{4M^2} \sum_{c} \sum_{v \in e} \sum_{w \in e} k_v k_w \quad (20) \]

Let

\[ d_c = \sum_{v \in c} k_v = \sum_{w \in c} k_w \quad (21) \]

Now can express modularity in terms of number of links and total degrees of nodes:

\[ Q = \frac{m_c}{M} - \frac{1}{4M^2} \sum_{c} \tilde{d}_c^2 \quad (22) \]

New modularity after \( n \) is removed from graph \( G \) can be written as:

\[ Q(G_n) = \frac{m_c - m_n}{M - k_n} - \frac{1}{4(M - k_n)^2} \sum_{c} \tilde{d}_c^2 \quad (23) \]

Last thing is to calculate the new \( \tilde{d}_c \). We can break this down in two cases:

**Case 1.** If \( c \neq c_n \) we have:

\[ \tilde{d}_c = \sum_{v \in c} k_v - m_n \quad (24) \]

**Case 2.** If \( c = c_n \) we have:

\[ \tilde{d}_c = \sum_{v \in c} k_v - m_n - k_n \quad (25) \]

Let:

\[ h_c = m_n + k_n \delta(c, c_n) \quad (26) \]

then finally we have:

\[ \tilde{d}_c = d_c - h_c \quad (27) \]

Giving us the final expression for the modularity once node \( n^* \) is removed:
\[ Q(G_n) = \frac{m_e - m_n}{M - k_n} - \frac{1}{4(M - k_n)^2} \sum_{c \in \varepsilon} (d_c - h_c)^2 \]  \hfill (28)

4 Experimental Results

As in [5], we evaluate performance using plots of the fraction of nodes removed, \( \rho \), and the fragmentation, \( \sigma \). Additionally, we plot the modularity as a function of \( \rho \). For each dataset, 8 methods of attack were tested: R, ID, RD, IB, RB, IM, RM, and MBA. We tested 4 datasets from the KONECT library: Arenas, Facebook, Powergrid, and Euroroad [7, 9, 10, 8]. The resulting plots are shown in Figures 1, 2, 3, and 4, respectively.

![Figure 1: Comparison of different network attack methods on arenas-meta network](image)

In all cases, the modularity-delta methods were the best at increasing the modularity of the networks. In terms of robustness, RB performs the best, as expected. We do see that our modularity-based
methods can perform just as well as MBA, and better than degree based attacks. Specifically, this is the case for the Powergrid Network and Euroroad network.
Figure 3: Comparison of different network attack methods on US powergrid network
Figure 4: Comparison of different network attack methods on euroroad network
5 Conclusion

In summary, we have shown that the calculation of modularity-deltas scales as $O(m)$, with a modularity-delta being the quantity modularity would change by if the node was deleted from a network. This quantity is useful in three key areas: community analysis, network robustness, and community detection. We have demonstrated that modularity-deltas are a useful semi-global centrality measure. When applied to network robustness, modularity-deltas were able to better fragment networks like the US Power Grid and the European road network, than other methods with similar complexity. Based on these results, we believe that modularity-deltas may further future research on robustness of large-scale social media networks.

References

[1] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
[2] Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.
[3] Jinyin Chen, Lihong Chen, Yixian Chen, Minghao Zhao, Shanqing Yu, Qi Xuan, and Xiaoniu Yang. Gai-based q-attack on community detection. *IEEE Transactions on Computational Social Systems*, 6(3):491–503, 2019.
[4] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical review E*, 70(6):066111, 2004.
[5] Bruno Requiao da Cunha, Juan Carlos González-Avella, and Sebastián Gonçalves. Fast fragmentation of networks using module-based attacks. *PloS one*, 10(11):e0142824, 2015.
[6] Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. Attack vulnerability of complex networks. *Physical review E*, 65(5):056109, 2002.
[7] KONECT. Caenorhabditis elegans network dataset, 2017.
[8] KONECT. Euroroad, 2017.
[9] KONECT. Facebook (nips) network dataset, 2017.
[10] KONECT. Us power grid, 2017.
[11] Mark EJ Newman. Modularity and community structure in networks. *Proceedings of the national academy of sciences*, 103(23):8577–8582, 2006.
[12] Vincent A Traag, Ludo Waltman, and Nees Jan van Eck. From louvain to leiden: guaranteeing well-connected communities. *Scientific reports*, 9(1):1–12, 2019.
[13] Stanley Wasserman, Katherine Faust, et al. *Social network analysis: Methods and applications*, volume 8. Cambridge university press, 1994.