A Methodology for Comparing and Creating Community Metrics

June Andrews
Center for Applied Math
Cornell University
Ithaca, NY 14853
http://www.cam.cornell.edu/ jandrews

John Hopcroft
Computer Science Dept
Cornell University
Ithaca, NY 14853
Email:jeh@cs.cornell.edu

***Abstract*—Community detection is used to understand the structure of large networks and their inner workings. It decomposes the network into a set of well-connected subgraphs, each of which is deemed a community. Complications arise in defining well-connected. Stemming from a variety of intuitions, several definitions have arisen. Two prominent definitions are: modularity [3], based off of an expected subgraph's density, and conductance[4], based off of the ratio of external edges to all edges of a subgraph. The communities determined by each definition are different and have only been compared through experimental methods.**

We present a framework that allows theoretical comparison of metric based community detection methods. We first parametrize metrics in terms of internal density and external sparsity of subgraphs. Once in this space, we use level sets to make clear what types of communities a metric will return. Our results support and expand upon published experimental results.

From this framework a new family of metrics becomes apparent. The Louvain Algorithm[2] is easily adapted to optimize for this new family of metrics and to return meaningful overlapping communities. We confirm our theoretical results with experiments on selected large networks and assert our improvement over modularity within our framework.

Keywords: Community Detection, Community Metrics, Networks

I. Introduction

Community detection allows us to decompose large networks into communities. The uses of these communities are for classifying nodes[]. A natural byproduct of decomposing large networks is to measure how well a decomposition describes a network. Applications, desire decompositions that select subsets of nodes with good internal connectivity. In a social graph, good connectivity of a subset of nodes indicates a group of friends, colleagues, or some common collective association between the people involved. In a citation network, good connectivity indicates a general topic shared between the papers.

The quality of a network decomposition in communities, or the strength of meaning of one network, have been analyzed based on metrics to measure their internal connectivity. Several metrics exist. The metrics are derived from intuitive definitions of what a community should be. Empirical comparisons of the metrics have been made. Empirical comparison to date have taken sets of networks, applied algorithms to maximize the metrics, and compared the communities on another criteria. In particular, Leskovec et al[1] compare detected communities with conductance, modularity ratio, and several other metrics. Leskovec et al reach conclusions about what types of communities each metric favors on given networks. We seek to expand the conclusions that can be reached by comparing metrics in a theoretical environment. We test the results on known data sets.

We present a simple framework to allow comparison of metrics before implementation. This framework reveals a methodology for creating new metrics, should an application demand. We provide a basic metric and test it theoretically and with specific implementations. We adapt the dendogram based on the Louvain algorithm [2], to quickly implement a metric maximization of these new metrics.

As applications for community metrics expand, we recommend this framework for adapting algorithms to application specific problems. Certain problems have more sensitivity or less space. Our framework allows customization to the needs of sensitivity to dense communities, encompassing all edges of the graph, and space to describe all communities.

This paper lays out community definitions for comparison of single communities and network covering community decompositions. Depending on the questions asked metrics for either case may be more appropriate. Our framework requires little adaptation between the two cases and produces new metrics for both. To start from a problem everyone is familiar with we use the Karate Club and Football Game network as our small scale network examples and the Physics Citation Archive as a larger scale network example. Maximizing the new metrics runs in time proportional to runtimes for modularity maximization. Hence, maximizing the new metric can be applied to larger networks.

Our notation throughout the paper will be conventional:

Network	$G(V, E)$
Nodes	$u, v \in V$
Weighted Edges	$w : V^2 \to \mathbb{R}_+ \leq 1$
Community	$C \subseteq V$
Size of C	$\|C\| = k\|V\|$
Set of Communities	$S = \{C_1, C_2, \ldots, C_n\}$

We use the convention that all edges are positively weighted, with a maximum weight of 1, and all possible edges exist. If

an edge does not exist in the network, we add it with 0 weight for ease of mathematical notation.

II. Metrics for Single Communities

Community definitions for single communities are metrics that determine whether or not a subset of node of the graph are appropriately well connected to each other and sparsely connected to the rest of the graph. We cover six popular metrics of single communities. There exist more definitions of single communities: (α, β), out degree fraction metrics, and variations of those. However, those definitions are not based on internal and external density, but rather on worst case analysis. We only consider the majority of definitions that are based on internal and external density of the entire community.

For a single community the accepted ideal community is a clique of nodes disconnected from the rest of the graph. This description depends on two parameters internal density and external sparsity. We formalize internal and external density.

Definition 1 (Internal Density):

$$I(C) = \frac{\sum_{u \in C} \sum_{v \in C} w(u, v)}{|C|(|C| - 1)}$$

Definition 2 (External Density):

$$E(C) = \frac{\sum_{u \in C} \sum_{v \notin C} w(u, v)}{|C|(|V| - |C|)}.$$

There are other representations of $I(C)$ and $E(C)$ that vary the $|C|$ and $|V| - |C|$ terms. The analysis and conclusions that follow are not sensitive those variations.. We can now dictate what the ideal community is mathematically:

Definition 3 (Ideal Single Community): A community, C, is ideal if it is an isolated clique and has the following properties:

$$I(C) = 1$$
$$E(C) = 0.$$

A. Specific Metrics

We cover six metrics for single communities:

- Conductance: the probability that a step in a random walk will leave the community [4].
- Cut Ratio: the fraction of existing to possible edges leaving the community [**?**].
- Edges Cuts: the number of edges connecting the community to the rest of the graph [**?**].
- Expansion: the average number of edges leaving the community per node [**?**].
- Internal Density: internal density of the community [**?**].
- Volume: the total degree of nodes within the community [**?**].

B. Parameterizations

We can now parameterize the described metrics in terms of internal and external density. We use the definitions from the previous section, the approximations $|C| \approx |C| - 1$, and the variable k to simplify $|C| = k|V|$. Hence we can put all previously described metrics in terms of $I(C)$, $E(C)$, $|C|$, $|V|$, and k.

$$\text{Conductance}(C) = \frac{(1-k)E(C)}{kI(C) + (1-k)E(C)}$$

$$\text{Cut Ratio} = k(1-k)|V|^2 E(C)$$

$$\text{Edge Cuts} = E(C)$$

$$\text{Expansion} = (1-k)|V|E(C)$$

$$\text{Internal Density} = 1 - I(C)$$

$$\text{Volume} = |C|^2 I(C) + (1-k)|C||V|E(C)$$

Our first observation is that not all metrics are functions of both $I(C)$ and $E(C)$. We address the implications of a metric defined over only $I(C)$ or $E(C)$ in later sections.

C. Greedy Optimization

We now use a simple greedy algorithm for each of these metrics to observe what types of communities optimize each metric. The algorithm is:

Algorithm 1 Greedy Single Community Metric Optimization

Input: C, $G = (V, E)$, and Metric
  $inc = 1$.
  while $inc \geq 0$ and $C \neq V$ **do**
    Let $u \in V$ maximize Metric$(C \cup u)$.
    $inc =$ Metric$(C \cup u) -$ Metric(C)
    Augment C by u
  end while

Some metrics require minimization rather than maximization, this algorithm is easily adapted accordingly.

For each metric we start the algorithm with a simple subset of two connected nodes and track the communities that result in improvements, according to a given metric. As the algorithm expands the community, the metric determines which node is optimal at each step. We record this better community and for each metric, M, get a set of $C_1, C_2, \ldots C_n$, such that the metric is increasing $M(C_i) < M(C_{i+1})$. While this is not an encompassing algorithm that checks for all communities, it will reveal a metric's biasenesses toward communities of certain $I(C)$, $E(C)$, values.

We apply the greedy algorithm to the well-known Zachary's Karate Club[] and the American College Football Game[] data

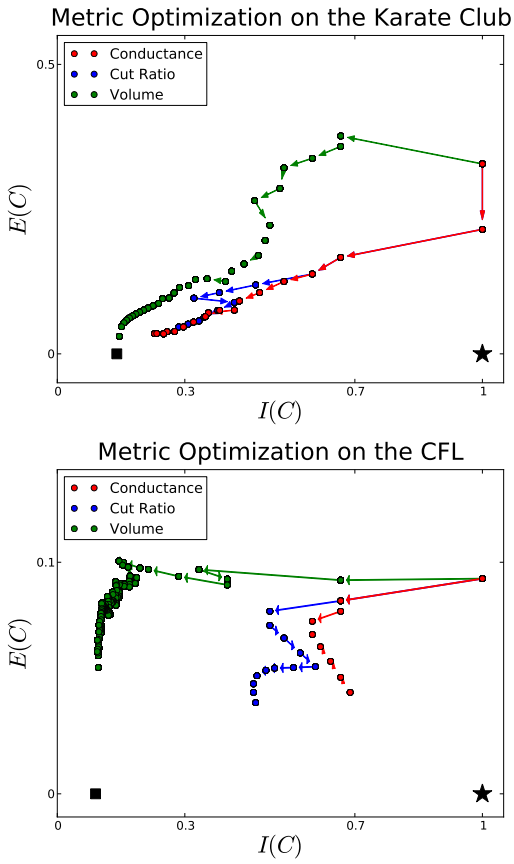

Fig. 1. Seeding the greedy algorithm with two connected nodes, we track the communities produced by optimizing each metric in the I, E plane. The black star is the definition of an ideal community and the black square is the I, E value of a community encompassing the entire graph.

sets and obtain *Fig. 1*. The metrics edge cuts, expansion, and internal density are not included as their paths overlaps other metric's paths. In the Karate Club optimizing conductance, cut ratio, and volume leads to nearly encompassing the entire graph. In the College Football Leagues the metric's behavior is a little less predictable.

D. Level Sets

So far we have parameterized the metrics in terms of internal and external density and shown that their optimization appears sporadic and suboptimal at times. However, the metrics are being optimized. What has not been taken into account is how the metrics optimize over the I, E plane. We can visualize their optimizations by calculating the value of the metric for all points of the I, E plane and graphing the contour curves, or level sets[]. In level set figures, any two points in the I, E plane connected by a curve have the same metric value. We now show how the metrics optimize in regard to the I, E plane and show that level sets explain the greedy optimization's path through the I, E plane.

1) Conductance Level Sets: For conductance the level sets are rays radiating from $(I, E) = (0, 0)$, see *Fig. 2*. As the rays become closer to horizontal, $E = 0$, conductance becomes optimal. Hence, no matter the data set, conductance desires communities with I, E values closer to $(I, E) = (x, 0)$, for any arbitrary $I(C) = x$. Including more nodes that continually increase $I(C)$ is rare in networks and much more common is decreasing $E(C)$ by encompassing more of the graph. Since, conductance is fairly, unresponsive to changes in $I(C)$, optimizing conductance results decreasing $E(C)$ as much as possible. How much conductance is biased towards small improvements in $E(C)$ verse large improvements in $I(C)$ depends on where in the I, E domain the seed communities originate. In the College Football League, the seed community begins a portion of the domain where improvements in I and E are balanced. For Zarchary's Karate Club, the seed community quickly falls into a portion of the I, E domain where small and easy improvements in E benefit conductance more than large and hard changes in I.

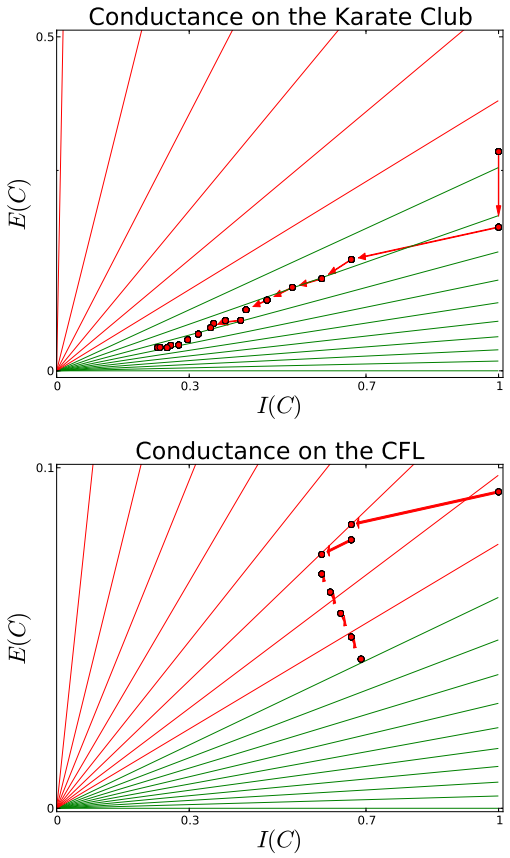

Fig. 2. *Conductance is optimized at* $E = 0$.The inclusion of level sets makes the movement of conductance across the I, E plane more comprehensible. The more level sets conductance crosses to make it towards $E = 0$, the better the conductance value of the communities is.

The last parameter that must be taken into account is k, the proportion of the graph covered by the community. Now we fix the I, E ratio and observe how changes in $|C| = k|V|$ affect conductance. The conclusion, If the expansion of the community is greatly affected by k, then our previous analysis will yield to conductance favoring the largest possible community as demonstrated in *Fig. 3*. This affect can be seen

in the change of C_1 to C_2 in the plotted paths in *Fig. 2*, as the relative size change of adding just one node is large.

That conductance favors large disconnected communities, even with low internal density, is experimentally confirmed by []. However, as long as the community is of medium size and has a much larger $I(C)$ value than $E(C)$ conductance will make improvements that correspond to our intuition that an ideal community is internal dense and externally sparse.

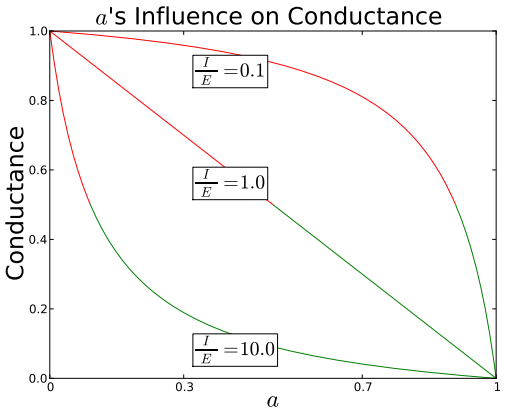

Fig. 3. *Conductance is optimized at* 0. The affect the size of the community has on CONDUCTANCE. Notice that the worst values of conductance correspond to small communities, irregardless of the I, E quality of the community. Similarly, the best conductance values correspond to the largest communities, irregardless of the internal and external density.

E. Cut Ratio, Edge Cuts, and Expansion

Similarly as for conductance, showing the level sets of cut ratio, edge cuts, and expansion across the I, E plane reveals the I, E values of the communities that optimize these metrics. While, the three metrics have different definitions, their level sets are visually identical only with different values. The level sets are horizontal lines, whose metric's are optimized by improvements in $E(C)$, irregardless of $I(C)$. Accordingly, the node that has the greatest improvement in $E(C)$ is the same for all three metrics, and each metric results in the same path through the I, E plane.

Though it does not make a difference in our examples, the difference between these metrics is their treatment of a. Cut ratio is unresponsive to changes in the size of the community, while expansion linearly discount against larger communities. Edge cuts heavily favors very large or very small communities. See *Fig. 5*

F. Internal Density

On the polar extreme of cut ratio, edge cuts, and expansion, is internal density. Internal density is a function of internal density and is unresponsive to changes in the the external density. We do not include indepth analysis, but rather a summary. The level sets of internal density are vertical lines in the I, E plane. The paths produced by the greedy algorithm grow out the original community to the largest clique it can find, as the internal density of two connected notes is already

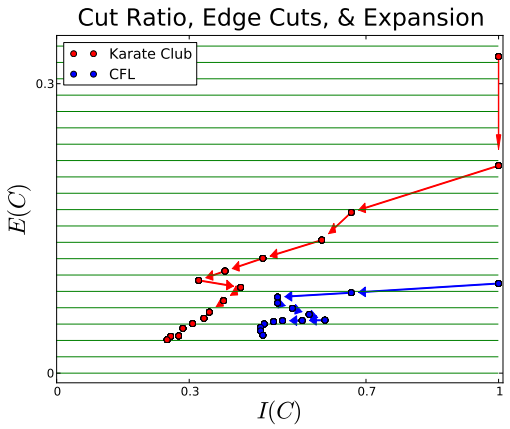

Fig. 4. *Metrics optimized at* $E(C) = 0$. Level sets show how cut ratio, edge cuts, and expansion evaluate the I, E plane. The path corresponds to the I, E values of communities expanded to optimize all three metrics. Because the level sets are the same, all three metrics produce the same path of optimizing communities in the Karate Club and CFL data sets.

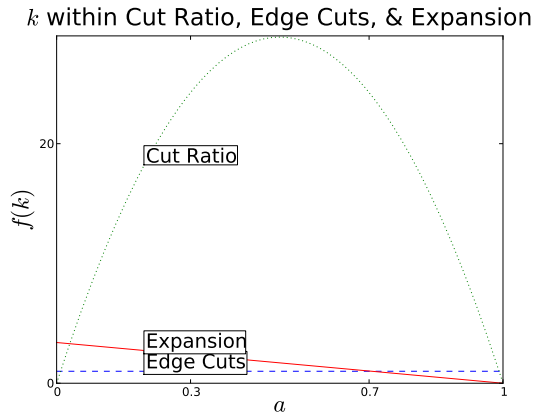

Fig. 5. Given a fixed $E(C)$, $f(k)$ is the constant multiple within each of the metrics, that incorporates a factor according to the size of $|C| = k|V$. These values are specific for the size of communities within the Karate Club.

maximal. Any clique is an optimization of internal density, irregardless of external connectivity.

G. Volume

A metric that takes both internal and external density into account is volume. The next conclusion is not apparent just from the equation parameterized in terms of internal and external density. However, observing the level sets of volume reveal that the optimal community is at $(I, E) = (0, 0)$ and volume as a metric is optimal for communities with low external density and low internal density. Apart from communities of disconnected nodes, volume can best be optimized by a community encompassing the entire graph. Volume directly contradicts our intuition that communities should have good internal connectivity.

H. Linearity

We maintain that the variables of internal and external density incorporate our full intuition of communities, metrics

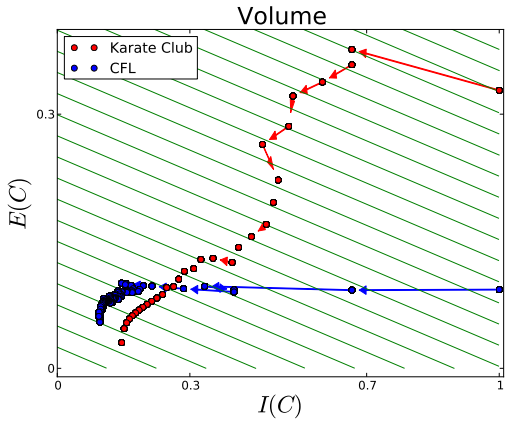

Fig. 6. Given a fixed $E(C)$, $f(a)$ is the constant multiple within each of the metrics, that incorporates a factor according to the size of $|C| = k|V|$. These values are specific for the size of communities within the Karate Club.

are just a matter of balancing between the two. We have shown how previous metrics single communities balance the two or select one to optimize for. We propose a linear metric that transparently balances between the two variables.

Definition 4 (Linearity): Our metric for single communities:

$$\text{LINEARITY} = M_L(C) = aI(C) - bE(C) \quad (1)$$

Many metrics incorporated the size of the community into their weighting. Indeed, it is more significant for a large community to gain nearly the same internal density as small community. Accordingly, there is a generalization of LINEARITY that can account for any desired weighting between internal and external density, with weighting for any size of community.

Definition 5 (General Metric): Our metric for single communities in its greatest generality:

$$\text{GENERAL} = M_G(C) = \sum_{i=0} f_i(C)I(C)^i - g_i(C)E(C)^i \quad (2)$$

Where f_i and g_i can be any function of the size of a community.

Linearity behaves similarly to conductance, when conductance is in a region of the I, E plane where there is a fair balance between improvements of I and E. Unlike conductance though, linearity does not have a critical point in the I, E plane, where the balance is shifted towards only favoring improvements in E. The result is that with the same initial seeds in the Karate Club and CFL, linearity follows the same path as conductance for the first few expansions and stops, rather than engulfing the network.

I. Optimal Communities

To conclude our analysis of metrics for the Karate Club and the CFL networks we present the optimal communities within the network according to each metric. For the Karate Club the optimal communities are:

Metric	Optimal C	$(I(C), E(C))$
CONDUCTANCE	G	$(0.14, 0)$
CUT RATIO	G	$(0.14, 0)$
EDGE CUTS	G	$(0.14, 0)$
EXPANSION	G	$(0.14, 0)$
INTERNAL DENSITY	$\{u, v\}\|w(u, v) = 1$	$(1, x)$
LINEARITY	clique of 3 nodes	$(1, 0.04)$
VOLUME	G	$(0.14, 0)$

The same trend holds for the CFL network, except with LINEARITY returning a set of different nodes according to a and b in the metric.

III. METRICS FOR SETS OF COMMUNITIES

Our original parameterization of internal and external density can not be directly applied to a set of communities, $S = \{C_1, C_2, \ldots C_n\}$. Though if we follow the same logic we will arrive at a similar parameterization. An ideal set of communities are cliques such that every edge is within some community. In addition, the community description of the network should not have an exponential number of communities, but rather some concise set of communities. Hence an ideal set of communities has three parameters. Internal density is a representation of how close the set of communities is to being a set of cliques. External density is a representation of how close the set of communities are to covering all edges in the graph. Size of the set of communities is a representation of how concise the set of communities are. We pick the following representations, but note that different representations do not yield different conclusions in the following sections.

Definition 6 (Internal Density for a Set of Communities):

$$I(S) = \sum_{C \in S} \frac{\sum_{u \in C} \sum_{v \in C} w(u, v)}{\sum_{C \in S} |C|(|C| - 1)}$$

Definition 7 (External Density for a Set of Communities):

$$E(S) = \frac{\sum_{(u,v) \in \text{EXT\_EDGES}} w(u, v)}{\sum_{u \in V} \sum_{v \in V} w(u, v)}$$

Definition 8 (Conciseness for a Set of Communities):

$$\text{CONCISENESS}(S) = |S|$$

To date we do not know of any previous metrics that have a closed form parameterization in terms of $I(S)$, $E(S)$, and $|S|$. We release a linear metric pertaining to these parameters in a later section.

Definition 9 (Ideal Set of Communities): A set of communities, S, is ideal if it is a set of cliques that cover the graph in the fewest necessary communities:

$$I(S) = 1$$
$$E(S) = 0$$
$$|S| = 1.$$

A. Modularity

Modularity is a partitioning of the network, and each partition is deemed a community. For each module its modularity is the difference of the existing internal edges and expected number of internal edges, had the graph been random, with the same distribution of node degrees. The total modularity is then the sum of the modularity of each partition. Modularity was developed by [] and has found wide spread use due to the fast algorithms for maximizing modularity [3]. In particular, the use of dendograms in the Louvain Algorithm [2] runs in minutes for large networks.

There is not a closed form parameterization of modularity in terms of our definitions of $I(S)$, $E(S)$, and $|S|$. But we find $I(S)$ and $E(S)$ to be intuitive enough to provide the space in which to analyze metrics of sets of communities.

Still, we would like to be able to use level sets for some visual explanation of how modularity behaves with regard to some internal and external density. The parameterization of modularity we can provide is in terms of $I(C)$ and $E(C)$ for single communities. If we allow, $p = \frac{|C|(|C|-1)}{2L}$ and $q = \frac{|C|(|V|-|C|)}{2L}$ to be constants:

$$\textsc{Modularity}(S) = \sum_{C \in S} pI(C) - \left(pI(C) + qE(C)\right)^2. \quad (3)$$

We can draw loose conclusions from this parameterization.

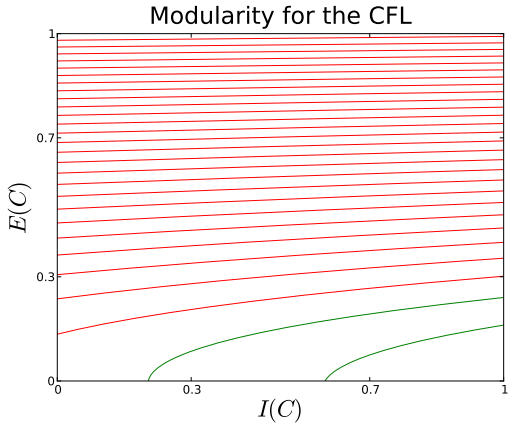

Fig. 7. The level sets of how MODULARITY treats the $I(C)$, $E(C)$ space for one community of size 9 in the CFL. Note the sharp transition from a region that heavily favors improvements in external density to a region that heavily favors improvements in internal density($E(C) < 0.1$).

The first check is that a set of disjoint cliques are the only communities that optimize modularity. There are no counter intuitive communities that optimize modularity. Overall, modularity is a two part optimization. For seed communities, in the case of modularity individual nodes, the typical initial values are high in external density. When $E(C)$ is high it is the dominating term and the optimization is to decrease $E(C)$ as quickly as possible. Once, external density is not the dominating term the optimization is then center around trying to improve $I(C)$. The transition between these two phases of optimization is sudden as attested by the level sets.

We would like to recommend modularity should be designed to optimize for improvements in internal density first to find the dense cores of communities and then optimize over external density to cover more of the graph. The reasoning is that improvements in external density can be made at any time by increasing the size of communities. However, improvements in internal density can only be made while the communities are small and closer to cliques.

B. Linearity for Sets of Communities

We now present a linear metric for sets of communities with regard to $I(S)$, $E(S)$, and $|S|$. It follows the same intuition of creating the linear metric for a single community. The ideal community is located at $(I(S), E(S), |S|) = (1, 0, 0)$ and accordingly the level sets are parallel planes emanating from around the ideal community. Hence, the linearity of the metric. How to balance between improvements in each of the parameters is set by the user.

Definition 10 (Linearity): Our metric for a set of communities:

$$\textsc{Linearity}(S) = aI(S) - bE(S) - c|S|,$$

where $a, b, c \geq 0$.

Depending on the application, communities of certain characteristics may be desired. Communities of size smaller than a certain size may be desired to be penalized or trade offs between improvements in $E(S)$ and $I(S)$ may depend on existing values. A general form of this equation is available.

Definition 11 (General Metric): Our metric for single communities in its greatest generality:

$$\textsc{General}(S) = \sum_{i=0} f_i(S)I(S)^i - g_i(S)E(S)^i - h_i|S|^i$$

C. Performance

We now empirically compare $\textsc{Linearity}(S)$ and $\textsc{Modularity}(S)$. Theoretically, modularity is expected to begin with trades of large improvements in external density, $E(S)$, for large losses of internal density, $I(S)$. Once the portion of graph is reached where internal density is more highly favored modularity will be maximized, as further improvements to internal density will not exist. Theoretically, linearity is expected to make the same trade off at every stage producing a nearly linear path through the $I(S)$, $E(S)$ space.

The compute the sets of communities that maximize each metric we use the Louvain algorithm[]. The algorithm does not need to be modified to maximize modularity. We easily adapt Louvain's algorithm to maximize linearity for sets of communities. Once we run the adapted Louvain algorithm for maximizing linearity we can further maximize linearity by expanding the partitions into overlapping communities. We note that no expansion making a full community a subset of another community will be optimal. Hence, we can use a greedy algorithm to expand the partitions produced by the Louvain algorithm. We expand each community to overlap neighboring communities if the metric is improved by doing so. We have tested running expansion off of the partition given

by maximizing modularity and find that the following results are still optimized by using the partition given by maximizing linearity. The following figures*Fig. 8*, *Fig.9*, and *Fig* show how the partitions produced by each metric being optimized are quickly divergent.

Maximizing the metrics involves a process of starting with set of communities, finding the maximal beneficial union of communities and iterating. Once a beneficial union is found the set of communities changes to reflect this. For each metric then we can produce a path of $S_1, S_2, \ldots$ that represent the intermediary sets of communities chosen by the metric. We can then plot this path of sets of communities in the $I(S), E(S)$ plane to see how the metric balances between internal and external density for a given network.

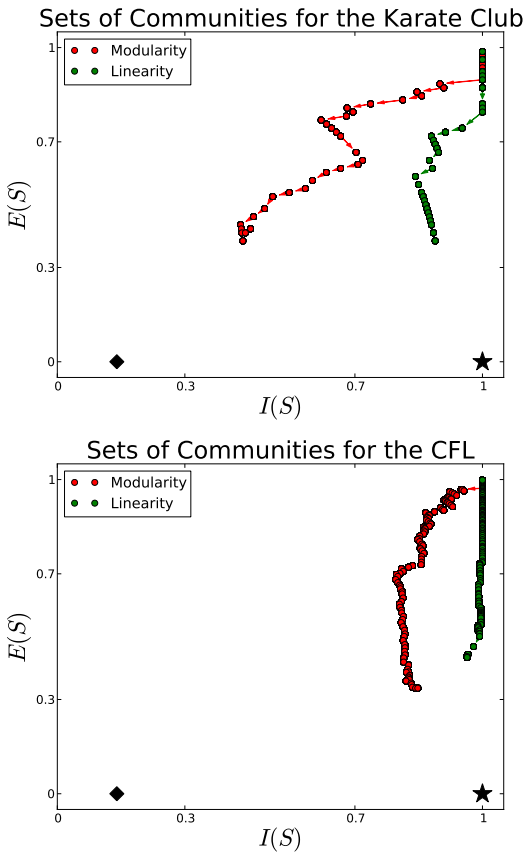

Fig. 8. The series of sets of communities produced by optimizing modularity and linearity, plotted by their $(I(S), E(S))$ values. The third parameter not displayed is how many communities are within each set. Linearity produces 16 and 18 communities for the karate club and the CFL respectively. Similarly. modularity produces 5 and 13 communities.

Because linearity is a direct optimization of parameters internal density, $I(S)$, external density $E(S)$, and conciseness, $|S|$, it can be modified through a, b, and c within LINEARITY(S) to always outperform MODULARITY(S) with regard to at least two parameters. In fact, it is only because of the inclusion of all three parameters that allows LINEARITY to return meaningful results. Trivial sets of communities exist for optimizing any two of the parameters. However, linearity

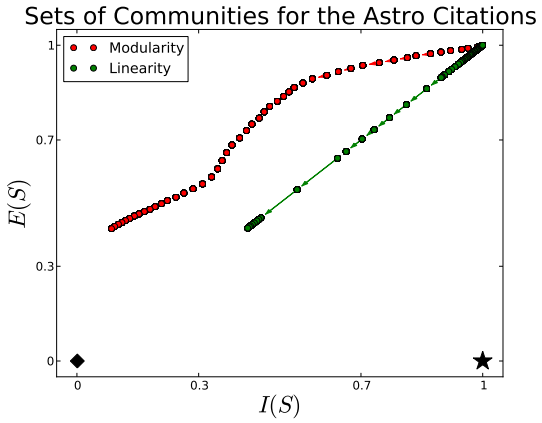

Fig. 9. Using the astrophysics citation network from [], with approximately 19k nodes and 200k edges, we apply the Louvain algorithm for modularity maximization and the adapted Louvain algorithm for linearity maximization. For the dimension not shown, conciseness, modularity produced 1523 communities and linearity produced 1258 communities. The points shown for linearity are an evenly spaced sample of changes to the set of communities. The points shown for modularity are the value of the communities at each stage of the dendogram. The black star corresponds to the ideal values for a set of communities. The black diamond corresponds to the value of a community encompassing the entire graph. Overall, linearity uses slightly fewer communities to have significantly higher internal density with comparable coverage of all edges of the graph.

exceeds expectations and outperforms in all three parameters for communities for the astrophysics citation network *Fig. 9*. If the partition produced by maximizing linearity is compared to the partition produced by maximizing modularity, in terms of modularity it is surprisingly close, considering that they are completely different optimizations with very different $I(S), E(S)$ values.

IV. CONCLUSION

The process we have used is straightforward. When designing a metric of testing the value of a single community or a set of communities, we began by describing the ideal communities. Once described the properties of the ideal communities could be named. In a single community's evaluation internal and external density are measured. A set of communities are evaluated on internal density, coverage of the network, and conciseness of the communities. Formalizing each of these properties, previous metrics could be analyzed in how they treated these properties. Visualization of their biasnesses could be seen from producing level sets. Different treatments of these properties then presented themselves. In particular, linear combinations of these parameters were revealed and transparently optimized towards the ideal values for internal and external density.

If dense internal structures and sparse external connectivity is desired in single communities, those properties can be directly optimized for with LINEARITY(C).

If dense internal structures, large coverage of the network, and a concise community description are desired, those properties can be directly optimized for with LINEARITY(S).

A. Future Work

While we have demonstrate a new framework and the metrics produced by an explicit relationship between internal density and external density, we have not fully explored the communities produced by varying that relationship. To date, we have only used iterative search of a, b, and c to find an optimal path through the internal and external density spaces. We have noticed that while there is infinite choice in setting a, b and c there are only a few possible paths through the I, E space, which has made finding appropriate values for a, b and c quick. We have not formalized this property or explored what it is about the structure of graphs that causes it.

The algorithm used to optimize linearity is a basic and fast algorithm, not designed around optimizing linearity. There is perhaps some gain to finding an algorithm that interweaves the step of finding larger communities and overlapping them. We have not tried the overlapping algorithm for modularity from [Ahn, etal].

More citations to follow.

Application to larger networks.

REFERENCES

[1] J. Leskovec, K. Lang, and M. Mahoney, *Empirical Comparison of Algorithms for Network Community Detection*, ACM WWW International conference on World Wide Web (WWW), 2010.

[2] V. Blondel, J. Guillaume, R. Lambiotte, and E. Lefebvre, *Fast Unfolding of Communities in Large Networks*, Journal of Statistical Mechanics: Theory and Experiment, 2008.

[3] M. Newman, *Modularity and community structure in networks*, PNAS, June 6, 2006 vol.103 no.23 8577-8582.

[4] B. Bollobas, *Modern Graph Theory*, Springer Verlag, New York, USA 1998.

[5] A. Clauset, M. Newman, and C. Moore, *Finding community structure in very large networks*, Physics Review E 70, 066111, 2004.