Karlsruher Institut für Technologie

Karlsruhe Reports in Informatics 2010,5
Edited by Karlsruhe Institute of Technology,
Faculty of Informatics
ISSN 2190-4782

Modularity-Driven Clustering of Dynamic Graphs

Robert Görke, Pascal Maillard,
Christian Staudt, Dorothea Wagner

2010

Fakultät für **Informatik**

Modularity-Driven Clustering of Dynamic Graphs*

Robert Görke[1], Pascal Maillard[2], Christian Staudt[1], and Dorothea Wagner[1]

[1] Institute of Theoretical Informatics
Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany
{rgoerke,christian.staudt,wagner}@ira.uka.de
[2] Laboratoire de Probabilités et Modèles Aléatoires
Université Pierre et Marie Curie (Paris VI), Paris, France
pascal.maillard@upmc.fr

Abstract. Maximizing the quality index *modularity* has become one of the primary methods for identifying the clustering structure within a graph. As contemporary networks are not static but evolve over time, traditional static approaches can be inappropriate for specific tasks. In this work we pioneer the NP-hard problem of online *dynamic modularity* maximization. We develop scalable dynamizations of the currently fastest and the most widespread static heuristics and engineer a heuristic dynamization of an optimal static algorithm. Our algorithms efficiently maintain a *modularity*-based clustering of a graph for which dynamic changes arrive as a stream. For our quickest heuristic we prove a tight bound on its number of operations. In an experimental evaluation on both a real-world dynamic network and on dynamic clustered random graphs, we show that the dynamic maintenance of a clustering of a changing graph yields higher *modularity* than recomputation, guarantees much smoother clustering dynamics and requires much lower runtimes. We conclude with giving sound recommendations for the choice of an algorithm.

1 Introduction

Graph clustering is concerned with identifying and analyzing the group structure of networks[3]. Generally, a partition (i.e., a clustering) of the set of nodes is sought, and the size of the partition is a priori unknown. A plethora of formalizations for what a *good* clustering is exist, good overviews are, e.g., [20, 2]. In this work we set our focus on the quality function *modularity*, coined by Girvan and Newman [3], which has proven itself feasible and reliable in practice, especially as a target function for maximization (see [1] for further references), which follows the paradigm of parameter-free community discovery [4].

The foothold of this work is that most networks in practice are not static. Iteratively clustering snapshots of a dynamic graph from scratch with a static method has several disadvantages: First, runtime cannot be neglected for large instances or environments where computing power is limited [5], even though very fast clustering methods have been proposed recently [6, 7]. Second, heuristics for the NP-hard [1] optimization of *modularity* suffer from local optima—this might be avoided by dynamically maintaining a good solution. Third, static heuristics are known not to react in a continuous way to small changes in a graph.

$$\begin{tikzcd} G \arrow[r, "\Delta"] \arrow[d, "\mathcal{T}"'] & G' \arrow[d, "\mathcal{T}"] \\ \mathcal{C}(G) \arrow[r, dashed, "\mathcal{A}"] & \mathcal{C}'(G') \end{tikzcd}$$

Fig. 1: The problem setting.

The lefthand figure illustrates the general situation for updating clusterings. A graph G is updated by some change Δ, yielding G'. We investigate procedures $\mathcal{A}$ that update the clustering $\mathcal{C}(G)$ to $\mathcal{C}'(G')$ without re-clustering from scratch, but work towards the same aim as a static technique $\mathcal{T}$ does.

*This work was partially supported by the DFG under grant WA 654/15-1.
[3]We use the terms *graph* and *network* interchangeably.

Related Work. Dynamic graph clustering has so far been a rather untrodden field. Recent efforts [8] yielded a method that can provably dynamically maintain a clustering that conforms to a specific bottleneck-quality requirement. Apart from that, there have been attempts to track communities over time and interpret their evolution, using static snapshots of the network, e.g. [9, 10], besides an array of case studies. In [11] a parameter-based dynamic graph clustering method is proposed which allows user exploration. Parameters are avoided in [12] where the minimum description length of a graph sequence is used to determine changes in clusterings and the number of clusters. In [13] an explicitly bicriterial approach for low-difference updates and a *partial* ILP[4] are proposed, the latter of which we also discuss. To the best of our knowledge no fast procedures for updating *modularity*-based clustering in general dynamic graphs have been proposed yet. Beyond graph theory, in data mining the issue of clustering an evolving data set has been addressed in, e.g., [14], where the authors share our goal of finding a smooth dynamic clustering. The literature on static *modularity*-maximization is quite broad. We omit a comprehensive review at this point and refer the reader to [1, 2, 15] for overviews, further references and comparisons to other clustering techniques. Spectral methods, e.g., [16], and techniques based on random walks [17, 18], do not lend themselves well to dynamization due to their non-continuous nature. Variants of greedy agglomeration [19, 6], however, are well suited, as we shall see.

Our Contribution. In this work we present, analyze and evaluate a number of concepts for efficiently updating *modularity*-driven clusterings. We prove the NP-hardness of dynamic *modularity* optimization and develop heuristic dynamizations of the most widespread [19] and the fastest [6] static algorithms, alongside apt strategies to determine the search space. For our fastest procedure, we can prove a tight bound of $\Theta(\log n)$ on the expected number of operations required. We then evaluate these and a heuristic dynamization of an ILP[4]-algorithm [1]. We compare the algorithms with their static counterparts and evaluate them experimentally on random preclustered dynamic graphs and on large real-world instances. Our results reveal that the dynamic maintenance of a clustering yields higher quality than recomputation, guarantees much smoother clustering dynamics and much lower runtimes.

Notation. Throughout this paper, we will use the notation of [20]. We assume that $G = (V, E, \omega)$ is an undirected, weighted, and simple graph[5] with the edge weight function $\omega \colon E \to \mathbb{R}_{\geq 0}$. We set $|V| =: n, |E| =: m$ and $\mathcal{C} = \{C_1, \ldots, C_k\}$ to be a partition of V. We call $\mathcal{C}$ a *clustering* of G and sets C_i *clusters*. $\mathcal{C}(v)$ is $C \ni v$. A clustering is *trivial* if either $k = 1$ ($\mathcal{C}^{\mathbf{1}}$), or all clusters contain only one element, i.e., are *singletons* ($\mathcal{C}^V$). We identify a cluster C_i with its node-induced subgraph of G. Then $E(\mathcal{C}) := \bigcup_{i=1}^{k} E(C_i)$ are *intra-cluster* edges and $E \setminus E(\mathcal{C})$ *inter-cluster* edges, with cardinalities $m(\mathcal{C})$ and $\overline{m}(\mathcal{C})$, respectively. Further, we generalize degree $\deg(v)$ to clusters as $\deg(C) := \sum_{v \in C} \deg(v)$. When using edge weights, all the above definitions generalize naturally by using $\omega(e)$ instead of 1 when counting edge e. Weighted node degrees are called $\omega(v)$. A *dynamic graph* $\mathcal{G} = (G_0, \ldots, G_{t_{\max}})$ is a sequence of graphs, with $G_t = (V_t, E_t, \omega_t)$ being the state of the dynamic graph at time step t. The *change* $\Delta(G_t, G_{t+1})$ between timesteps comprises a sequence of b *atomic* events on G_t, which we detail later (see Tab. 1). In our setting the sequence of changes arrives as a stream.

[4]ILP stands for Integer Linear Program

[5]A *simple* graph in this work is both loopless and has no parallel edges.

The Quality Index *Modularity*. In this work we set our focus on *modularity* [3], a measure for the goodness of a clustering. Just like any other quality index for clusterings (see, e.g., [20, 2]), *modularity* does have certain drawbacks such as *non-locality* and *scaling behavior* [1] or *resolution limit* [21]. However, being aware of these peculiarities, *modularity* can very well be considered a robust and useful measure that closely agrees with intuition on a wide range of real-world graphs, as observed by myriad studies. *Modularity* can be formulated as

$$\operatorname{mod}(\mathcal{C}) := \frac{m(\mathcal{C})}{m} - \frac{1}{4m^2} \sum_{C \in \mathcal{C}} \left(\sum_{v \in C} \deg(v) \right)^2 \quad \text{(weighted version analogous)} \ . \tag{1}$$

Roughly speaking, *modularity* measures the fraction of edges which are covered by a clustering and compares this value to its expected value, given a random rewiring of the edges which, on average, respects node degrees. This definition generalizes in a natural way as to take edge weights $\omega(e)$ into account, for a discussion thereof see [22] and [23]. ModOpt, the problem of optimizing *modularity* is NP-hard [1], but *modularity* can be computed in linear time and lends itself to a number of simple greedy maximization strategies. For the dynamic setting, the following simple corollary theoretically corroborates the use of heuristics, even if we do take the effort to compute an optimal initial clustering.

Corollary 1 **(DynModOpt is NP-hard).** *Given graph G, a modularity-optimal clustering $\mathcal{C}^{\text{opt}}(G)$ and an atomic event Δ to G, yielding G'. It is NP-hard to find a modularity-optimal clustering $\mathcal{C}^{\text{opt}}(G')$.*

Proof. We reduce an instance G of ModOpt to a linear number of instances of DynModOpt. Given graph G, there is a sequence $\mathcal{G}$ of graphs $(G_0, \ldots, G_\ell = G)$ of linear length such that (i) $\mathcal{G}$ starts with G_0 consisting of one edge e of G and its incident nodes u, v, (ii) $\mathcal{G}$ ends with G, (iii) graph G_{i+1} results from G_i and an atomic event Δ_i. ModOpt can be solved in constant time for G_0 yielding $\mathcal{C}^{\text{opt}}(G_0)$. Subsequently solving DynModOpt for instances $G_i, \mathcal{C}^{\text{opt}}(G_i), \Delta_i$ yielding $\mathcal{C}^{\text{opt}}(G_{i+1})$, we end with $\mathcal{C}^{\text{opt}}(G_\ell) = \mathcal{C}^{\text{opt}}(G)$, the solution to ModOpt.

Measuring the Smoothness of a Dynamic Clustering. By comparing consecutive clusterings, we quantify how smooth an algorithm manages the transition between two steps, an aspect which is crucial to both readability and applicability. An array of measures exist that quantify the (dis)similarity between two partitions of a set; for an overview and further references, see [24]. Our results strongly suggest that most of these widely accepted measures are qualitatively equivalent in all our (non-pathological) instances (see App. E for an example). We thus restrict our view to the *(graph-structural) Rand* index [24], being a well known representative; it maps two clusterings into the interval $[0, 1]$, i.e., from equality to maximum dissimilarity: $\mathcal{R}_g(\mathcal{C}, \mathcal{C}') := 1 - (|E_{11}| + |E_{00}|)/m$, with $E_{11} = \{\{v, w\} \in E : \mathcal{C}(v) = \mathcal{C}(w) \wedge \mathcal{C}'(v) = \mathcal{C}'(w)\}\}$, and E_{00} the analog for inequality. We use the intersection of two graphs when comparing their clusterings. *Low* distances correspond to *smooth* dynamics.

2 The Clustering Algorithms

Formally, a dynamic clustering algorithm is a procedure which, given the previous state of a dynamic graph G_{t-1}, a sequence of graph events $\Delta(G_{t-1}, G_t)$ and a clustering $\mathcal{C}(G_{t-1})$ of the previous state, returns a clustering $\mathcal{C}'(G_t)$ of the current state. While the algorithm may discard $\mathcal{C}(G_{t-1})$ and simply start from scratch, a good dynamic algorithm will harness the results of its previous work. A natural approach to dynamizing an agglomerative clustering algorithm is to break up those local parts of its previous clustering, which are most likely to require a reassessment after some changes to the graph. The half finished instance is then given to the agglomerative algorithm for completion. A crucial ingredient thus is a *prep strategy* S which decides on the search space which is to be reassessed. We will discuss such strategies later, until then we simply assume that S breaks up a reasonable part of $\mathcal{C}(G_{t-1})$, yielding $\tilde{\mathcal{C}}(G_{t-1})$ (or $\tilde{\mathcal{C}}(G_t)$ if including the changes in the graph itself). We call $\tilde{\mathcal{C}}$ the *preclustering* and nodes that are chosen for individual reassessment *free* (can be viewed as singletons).

Formalization of Graph Events. We describe our test instances in more detail later, but for a proper description of our algorithms, we now briefly formalize the graph events we distinguish. Table 1 lists all events (and their nomenclature) that can make up the sequence of changes between two graph states. Most commonly *edge creations* and *removals* take place, and they require the incident nodes to be present before and after the event. Given edge *weights*, changes require an edge's presence. *Node creations* and *removals* in turn only handle degree zero nodes, i.e., for an intuitive node deletion we first have to remove all incident edges. To summarize such compound events we use *time step* events, which indicate to an algorithm that an updated clustering must now be supplied. Between time steps it is up to the algorithm how it maintains its intermediate clustering. Additionally, *batch* updates allow for only running an algorithm after a scalable number of b timesteps.

Table 1: Atomic graph events

name	*formal*
node creation	$V + u$
node removal	$V - u$
edge creation	$E + \{u, v\}$
edge removal	$E - \{u, v\}$
weight increase	$\omega(u, v) + x$
weight decrease	$\omega(u, v) - x$
time step	$t + 1$

2.1 Algorithms for Dynamic Updates of Clusterings

The Global Greedy Algorithm. The most prominent algorithm for *modularity* maximization is a global greedy algorithm [19], which we call Global (Alg. 1). Starting with singletons, for each pair of clusters, it determines the increase in *modularity* dQ that can be achieved by merging the pair and performs the most beneficial merge. This is repeated until no more improvement is possible. As the pseudo-dynamic algorithm sGlobal[6], we let this algorithm cluster from scratch at each timestep for comparison. By passing a *preclustering* $\tilde{\mathcal{C}}(G_t)$ to Global we can define the properly dynamic algorithm dGlobal. Starting from $\tilde{\mathcal{C}}(G_t)$ this algorithm lets Global perform greedy agglomerations of clusters.

Algorithm 1: Global$(G, \mathcal{C})$

```
1 while $\exists C_i, C_j \in \mathcal{C} : \mathrm{dQ}(C_i, C_j) \geq 0$ do
2     $(C_1, C_2) \leftarrow \arg\max_{C_i, C_j \in \mathcal{C}} \mathrm{dQ}(C_i, C_j)$
3     merge$(C_1, C_2)$
```

The Local Greedy Algorithm. In a recent work [6] the simple mechanism of the aforementioned Global has been modified as to rely on local decisions (in terms of graph locality), yielding an extremely fast and efficient maximization. Instead of looking globally for the best merge of two clusters, Local, as sketched out in Alg. 2, repeatedly lets each node consider moving to one of its neighbors' clusters, if this improves *modularity*; this potentially merges clusters, especially when starting with singletons. As soon as no more nodes move, the current

[6]For historical reasons, sGlobal appears in plots as StaticNewman, dGlobal as Newman, sLocal as StaticBlondel and dLocal as Blondel, based on the algorithms' authors.

clustering is *contracted*, i.e., each cluster is contracted to a single node, and adjacencies and edge weights between them are summarized. Then, the process is repeated on the resulting graph which constitutes a higher level of abstraction; in the end, the highest level clustering is decisive about the returned clustering: The operation unfurl assigns each elementary node to a cluster represented by the highest level cluster it is contained in.

We again sketch out an algorithm which serves as the core for both a static and a dynamic variant of this approach, as shown in Algorithm 2. As the input, this algorithm takes a hierarchy of graphs and clusterings and a search space policy P. Policy P affects the graph *contractions*, in that P decides which nodes of the next level graph should be free to move. Note that the input hierarchy can also be flat, i.e., $h_{\max} = 0$, then line 11 creates all necessary higher levels.

Algorithm 2: Local($G^{0 \dots h_{\max}}, \mathcal{C}^{0 \dots h_{\max}}, P$)

```
1  $h \leftarrow 0$
2  repeat
3      $(G, \mathcal{C}) \leftarrow (G^h, \mathcal{C}^h)$
4      repeat
5          forall free $v \in V$ do
6              if $\max_{v \in N(u)} \mathrm{dQ}(u, v) \geq 0$ then
7                  $w \leftarrow \arg\max_{v \in N(u)} \mathrm{dQ}(u, v)$
8                  move($u, \mathcal{C}(w)$)
9      until no more changes
10     $\mathcal{C}^h \leftarrow \mathcal{C}$
11     $(G^{h+1}, \tilde{\mathcal{C}}^{h+1}) \leftarrow$ contract($G^h, \mathcal{C}^h, P$)
12     $h \leftarrow h + 1$
13 until no more real contractions
14 $\mathcal{C}(G^0) \leftarrow$ unfurl($\mathcal{C}^{h-1}$)
```

Again posing as a pseudo-dynamic algorithm, the static variant (as in [6]), sLocal, passes only $(G_t, \tilde{\mathcal{C}}^V)$ to Local, such that it starts with singletons and all nodes freed, instead of a proper *preclustering*. The policy P is set to tell the algorithm to also start from scratch on all higher levels and to not work on previous results in line 11, i.e., in $\tilde{\mathcal{C}}^{h+1}$ again all nodes in the contracted graph are free singletons.

The dynamic variant dLocal remembers its old results. It passes the changed graph, a current *preclustering* of it and all higher-level contracted structures from its previous run to Local: (G_t, $G_{\text{old}}^{1,\dots,h_{\max}}$, $\tilde{\mathcal{C}}$, $\mathcal{C}_{\text{old}}^{1,\dots,h_{\max}}$, P). In level 0, the preclustering $\tilde{\mathcal{C}}$ defines the set of free nodes. In levels beyond 0, policy P is set to have the contract-procedure free only those nodes of the next level, that have been affected by lower level changes (or their neighbors as well, tunable by policy P). Roughly speaking, dLocal starts by letting all free (elementary) nodes reconsider their cluster. Then it lets all those (super-)nodes on higher levels reconsider their cluster, whose content has changed due to lower level revisions. Thus, a run of Algorithm 2 avoids recomputing unrelated regions of the graph and resolving ambiguous situations in a complementary fashion without necessity.

ILP. While optimality is out of reach, the problem *can* be cast as an ILP [1]. A distance relation indicates whether elements are in the same cluster:

$$\forall \{u, v\} \in \binom{V}{2} : X_{uv} = \begin{cases} 0 & \text{if } \mathcal{C}(u) = \mathcal{C}(v) \\ 1 & \text{otherwise} \end{cases} \qquad (\text{ node distance var.}) \tag{2}$$

$$\forall \{u, v, w\} \in \binom{V}{3} : \begin{cases} X_{uv} + X_{vw} - X_{uw} \geq 0 \\ X_{uv} + X_{uw} - X_{vw} \geq 0 \\ X_{uw} + X_{vw} - X_{uv} \geq 0 \end{cases} ; \quad X_{uv} \in \{0, 1\} \tag{3}$$

$$\text{minimize } \mathrm{mod}_{\text{ILP}}(G, \mathcal{C}_G) = \sum_{\{u,v\} \in \binom{V}{2}} \left(\omega(u, v) - \frac{\omega(u) \cdot \omega(v)}{2 \cdot \omega(E)} \right) X_{uv} \tag{4}$$

Table 3: Reactions of the algorithms to graph events. Isolated nodes are made singletons when inserted and simply deleted when removed. With "$\rightarrow S$" we indicate that a *prep strategy* prepares a *preclustering*.

$\Delta(G_t, G_{t+1})$	Algorithms' reactions					
formal	sGlobal	dGlobal	sLocal	dLocal	dILP	dEOO
$E + \{u, v\}$	–	$\rightarrow S$	–	$\rightarrow S$	$\rightarrow S$, pILP(args)	-
$E - \{u, v\}$	–	$\rightarrow S$	–	$\rightarrow S$	$\rightarrow S$, pILP(args)	-
$\omega(u, v) + x$	–	$\rightarrow S$	–	$\rightarrow S$	$\rightarrow S$, pILP(args)	-
$\omega(u, v) - x$	–	$\rightarrow S$	–	$\rightarrow S$	$\rightarrow S$, pILP(args)	-
$t + 1$	Global $(G_t, \mathcal{C}^V)$	Global $(G_t, \tilde{\mathcal{C}})$	Local(G_t $\mathcal{C}^{\mathbf{1}}$,all)	Local($G_{t-1}^{1\ldots h_{\max}}$, $\tilde{\mathcal{C}}, \mathcal{C}_{t-1}^{1\ldots h_{\max}}$, aff/nb)	-	EOO(G_{t+1}, $\mathcal{C}_{t+1}$,args)

Note that the definition of X_{uv} renders this a minimization problem. Since runtimes for the full ILP reach days for more than 200 nodes, a promising idea pioneered in [13] is to solve a *partial ILP* (pILP). Such a program takes a *preclustering*—of much smaller complexity—as the input, and solves this instance, i.e., finishes the clustering, optimally via an ILP; a singleton *preclustering* yields a full ILP. We introduce two variants, (i) the argument noMerge prohibits merging *pre-clusters*, and only allows free nodes to join clusters or form new ones, and (ii) merge allows existing clusters to merge. For both variants we need to add constraints and terms to Eqs. 2-4. Roughly speaking, for (i), variables Y_{uC} indicating the distance of node u to cluster C are introduced and triplets of constraints similar to Eq. 3 ensure their consistency with the X-variables; for (ii), we additionally need variables $Z_{CC'}$ for the distance between clusters, constrained just as in Eq. 3. Details on these formulations and on their impact on the target function can be found in App. B. The dynamic clustering algorithms which first solicit a *preclustering* and then call pILP are called dILP. Note that they react on any edge event; accumulating events until a timestep occurs can result in prohibitive runtimes.

Elemental Optimizer The *elemental operations optimizer*, EOO, performs a limited number of operations, trying to increase the quality. Specifically, we allow moving or splitting off nodes and merging clusters, as listed in Table 2. Although rather limited in its options, EOO or very similar tools for local optimization are often used as post-processing tools (see [25] for a discussion). Our algorithm dEOO simply calls EOO at each time step.

Table 2: EOO operations, allowed/disallowed via parameters

Operation	Effect
merge(u,v)	$\mathcal{C}(u) \cup \mathcal{C}(v)$
shift(u,v)	$\mathcal{C}(u) - u, \mathcal{C}(v) + u$
split(u)	$(\{u\}, \mathcal{C}(u) \setminus u) \leftarrow \mathcal{C}(u)$

2.2 Strategies for Building the Preclustering

We now describe *prep strategies* which generate a *preclustering* $\tilde{\mathcal{C}}$, i.e., define the search space. We distinguish the *backtrack strategy*, which refines a clustering, and *subset strategies*, which free nodes. The rationale behind the *backtrack strategy* is that selectively backtracking the clustering produced by Global enables it to respect changes to the graph. On the other hand, *subset strategies* are based on the assumption that the effect of a change on the clustering structure is necessarily local. Both output a half-finished *preclustering*.

The *backtrack strategy* (BT) records the merge operations of Global and backtracks them if a graph modification suggests their reconsideration. We detail in App. C what we mean by "suggests", but for brevity we just state that the actions listed for BT provably require very little asymptotic effort and offer Global a good chance to find an improvement. Speaking intuitively, the reactions to a change in (non-)edge $\{u, v\}$ are as follows (weight changes are analogous): For intra-cluster additions we backtrack those merge operations that led to u and

Table 4: Overview of how strategies handle graph events. Changes to edges' weights are analog to creations/removals. Degree-0 nodes are universally made singletons when inserted and ignored when removed.

Event	Reaction			
	BT	BU, $\tilde{V} =$	N, $\tilde{V} =$	BN, $\tilde{V} =$
$E + \{u, v\}$	$\begin{cases} \mathsf{sep}(u,v) & \mathcal{C}(u) = \mathcal{C}(v) \\ \mathsf{iso}(u), \mathsf{iso}(v) & \mathcal{C}(u) \neq \mathcal{C}(v) \end{cases}$	$\mathcal{C}(u) \cup \mathcal{C}(v)$	$N_d(u) \cup N_d(v)$	$\mathrm{BFS}\{u, v\}_{\vert s}$
$E - \{u, v\}$	$\begin{cases} \mathsf{iso}(u), \mathsf{iso}(v) & \mathcal{C}(u) = \mathcal{C}(v) \\ - & \mathcal{C}(u) \neq \mathcal{C}(v) \end{cases}$	$\mathcal{C}(u) \cup \mathcal{C}(v)$	$N_d(u) \cup N_d(v)$	$\mathrm{BFS}\{u, v\}_{\vert s}$

v being in the same cluster and allow Global to find a tighter cluster for them, i.e., we separate them. For inter-cluster additions we track back u and v individually, until we isolate them as singletons, such that Global can re-classify and potentially merge them. Inter-cluster deletions are not reacted on. On intra-cluster deletions we again isolate both u and v such that Global may have them find separate clusters. For more details on these operations see App. C. Note that this strategy is only applicable to Global; conferring it to Local is neither straightforward nor promising as Local is based on node *migrations* in addition to *agglomerations*. Anticipating this strategy's low runtime, we can give a bound on the expected number of backtrack steps for a single call of the crucial operation isolate. We leave its formal proof to The. 2 in App. C:

Theorem 1. *Assume that a backtrack step divides a cluster randomly. Then, for the number I of steps isolate(v) requires, it holds: $E\{I\} \in \Theta(\ln n)$.*

A *subset strategy* is applicable to all dynamic algorithms. It frees a subset $\tilde{V}$ of individual nodes that need reassessment and extracts them from their clusters. We distinguish three variants which are all based on the hypothesis that local reactions to graph changes are appropriate. Consider an edge event involving $\{u, v\}$. The *breakup strategy* (BU) marks the affected clusters $\tilde{V} = \mathcal{C}(u) \cup \mathcal{C}(v)$; the *neighborhood strategy* (N_d) with parameter d marks $\tilde{V} = N_d(u) \cup N_d(v)$, where $N_d(w)$ is the d-hop neighborhood of w; the *bounded neighborhood strategy* (BN_s) with parameter s marks the first s nodes found by a breadth-first search simultaneously starting from u and v.

3 Experimental Evaluation of Dynamic Algorithms[7]

Instances. We use both generated graphs and real-world instances. We briefly describe them here, but for more details please see [26] and [13].

Random Graphs $\{$*ran*$\}$. Our Erdős-Rényi-type generator builds upon [27] and adds to this dynamicity in all graph elements and in the clustering, i.e., nodes and edges are inserted and removed and ground-truth clusters merged and split, always complying with sound probabilities. The generator's ground-truth clustering defines edge probabilities and thereby steers how the graph evolves. Roughly speaking, within ground-truth clusters edges accumulate and in between they become sparse. The generator additionally maintains a reference clustering, which follows the ground-truth clustering, as soon as changes in the latter actually manifest in the edge structure; we use this reference clustering to compare our algorithms to. In later plots we use selected random instances, however, descriptions apply to all such graphs.[7]

EMail Graph $\mathcal{G}_e$. The network of email contacts at the department of computer science at KIT is an ever-changing graph with an inherent clustering: Workgroups and projects cause

[7]For implementation notes see App. A, supplementary information is stored at i11www.iti.uni-karlsruhe.de/projects/spp1307/dyneval

increased communication. We weigh edges by the number of exchanged emails during the past seven days, thus edges can completely time out; degree-0 nodes are removed from the network. This network, $\mathcal{G}_e$, has between 100 and 1500 nodes depending on the time of year, and about 700K events spanning about 2.5 years. It features a strong power-law degree distribution. Fig. 2 shows the temporal development of this graph in terms of n (**lower**) and m (**upper**) per 100 events. The first peak stems from a spam attack in late '06, the two large drops from Christmas breaks and the smaller drops from spring and autumn breaks.

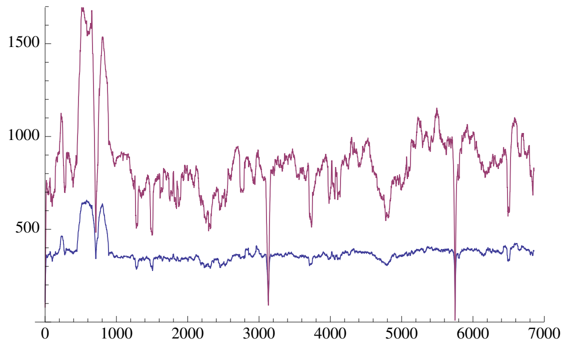

Fig. 2: Nodes (**blue**) and edges (**red**) of $\mathcal{G}_e$

arXiv Graphs {arx}. Since 1992 the *arXiv.org e-Print archive*[8] is a popular repository for scientific e-prints, stored in several categories alongside timestamped metadata. We extracted networks of collaboration between scientists based on coauthorship. For each e-print we add equally weighted clique-edges among the contributors such that each author gains a total edge weight of 1.0 per e-print contributed to; see Fig. 3 for three examples. We let e-prints time out after two years and remove disconnected authors. As these networks are ill-natured for local updates, we shall use them as tough trials. We show results for two categories that feature a large connected component.

Fundamental Results. For the sake of readability, we use a moving average in plots for distance and quality to smoothen the raw data (see Figs. 18 and 19 in App. E for an example), separately looking at variances. We consider the criteria quality (*modularity*), smoothness ($\mathcal{R}_g$) and runtime (ms), and additionally $|\mathcal{C}|$.

Discarding dEOO. In a first feasibility test, dEOO immediately falls behind all other algorithms in terms of quality (Fig. 10), an observation substantiated by the fact that postprocessing works better if related to the underlying algorithm [25]. Moreover, runtimes for dEOO as the sole technique are infeasible for large graphs.

Local Parameters. It has been stated in [6] that the order in which Local considers nodes is irrelevant. In terms of average runtime and quality we can confirm this for sLocal, though a random order tends to be less smooth; for dLocal the same observation holds (see App. D.3). However, since node order *does* influence specific values, a random order can compensate the

[8]Website of e-print repository: arxiv.org; for our tools for collecting and processing the data see i11www.iti.uni-karlsruhe.de/projects/spp1307/dyneval

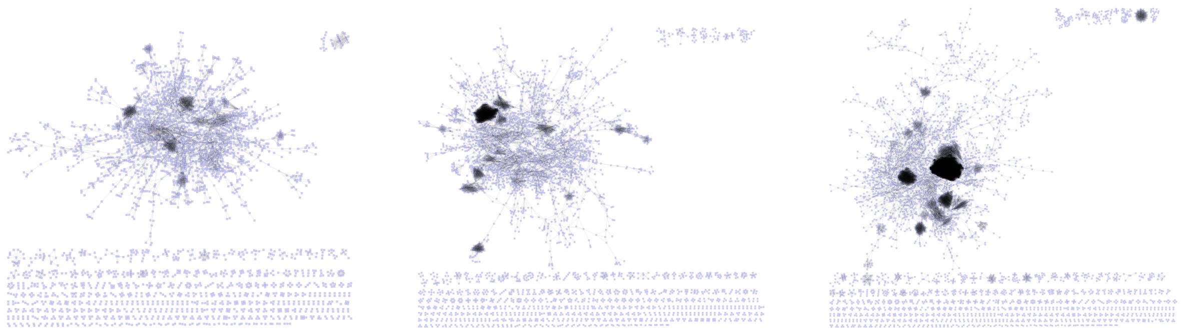

Fig. 3: The *arXiv* category *Nuclear Theory*, containing 33k e-prints, after '01, '05 and '09. Since disconnected cliques let *modularity* approach 1 quickly, lower *modularity* corresponds to a higher degree of collaboration.

effects this might have in pathological cases. We found that considering only affected nodes or also their neighbors in higher levels, does not affect any criterion on average.

pILP Variants. Allowing the ILP to merge existing clusters takes longer, and clusters coarser and with a slightly worse *modularity*; we therefore reject it.

Heuristics vs. dILP. A striking observation about dILP is the fact that it yields worse quality than dLocal and sLocal with identical *prep strategies*, as in Fig. 4. Being locally optimal seems to overfit, a phenomenon that does not weaken over time and persists throughout most instances. Together with its high runtime and only small advantages in smoothness dILP is ill-suited for updates on large graphs.

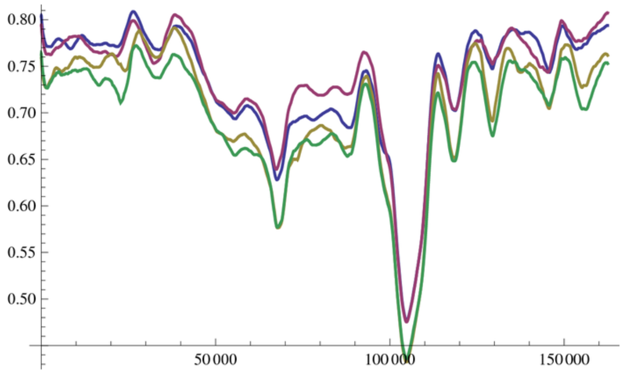

Fig. 4: *Modularity* for dLocal (**blue**), dGlobal (**purple**), dILP(noMerge) (**yellow**) and dILP(merge) (**green**) on the first quarter of $\mathcal{G}_e$, batch size 1, strategy BN_8

Static Algorithms. Briefly comparing sGlobal and sLocal we can state that sLocal consistently yields better quality and a finer yet less smooth clustering (see App. D.2). This generally applies to the corresponding dynamic algorithms as well. In terms of speed, however, sGlobal hardly lags behind sLocal, especially for small graphs with many connected components, where sLocal cannot capitalize on its strength of quickly reducing the size of a large instance. For such instances, separately maintaining and handling connected components could thus reasonably speed up sLocal, but would also do so for sGlobal.

Prep Strategies. We now determine the best choice of *prep strategies* and their parameters for dGlobal and dLocal. In particular, we evaluate N_d for $d \in \{0, 1, 2, 3\}$ and BN_s for $s \in \{2, 4, 8, 16, 32\}$, alongside BU and BT. Throughout our experiments $d = 0$ (or $s = 2$) proved insufficient, and is therefore ignored in the following. For dLocal, increasing d has only a marginal effect on quality and smoothness, while runtime grows sublinearly, which suggests $d = 1$. For dGlobal, N_d risks high runtimes for depths $d > 1$, especially for dense graphs. In terms of quality N_1 is the best choice, higher depths seem to deteriorate quality—a strong indication that large search spaces contain local optima. Smoothness approaches the bad values of sGlobal for $d > 2$. For BN, increasing s is essentially equivalent to increasing d, only on a finer scale. Consequently, we can report similar observations. For dLocal, BN_4 proved slightly superior. dGlobal's quality benefits from increasing s in this range (see App. E.1), but again at the cost of speed and smoothness, so that BN_{16} is a reasonable choice. BU clearly falls behind in terms of all criteria compared to the other strategies, and often mimics the static algorithms. dGlobal using BT is by far the fastest algorithm, confirming our theoretical predictions from Sec. 2.2, but still produces competitive quality. However, it often yields a smoothness in the range of sGlobal. Summarizing, our best dynamic candidates are the algorithms dGlobal@BT and dGlobal@BN_{16} (achieving a speedup over sGlobal of up to 1k and 20 at 1k nodes, respectively) and algorithm dLocal@BN_4(speedup of 5 over sLocal).

Comparison of the Best. As a general observation, as depicted in Fig. 6, each dynamic candidate beats its static counterpart in terms of *modularity*. On the generated graphs, dLocal

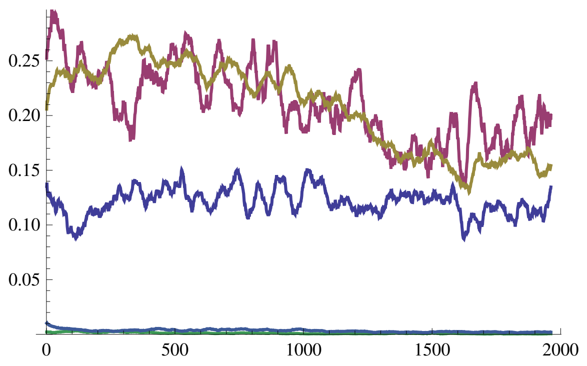

Fig. 5: $\mathcal{R}_g$, {ran}: sGlobal and sLocal are less smooth (factor 100) than dLocal@BN$_4$, dGlobal@BN$_{16}$ (bottom); dGlobal@BT competes well.

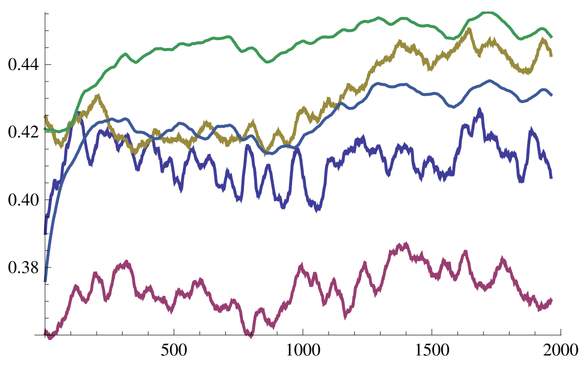

Fig. 6: *Modularity*, {ran}: dGlobal@BT (lower blue) and dGlobal@BN$_{16}$ (upper blue) beat sGlobal; dLocal@BN$_4$ beats sLocal.

is superior to dGlobal, and faster. In terms of smoothness (Fig. 5), dynamics (except for dGlobal@BT) are superior to statics by a factor of ca. 100, but even dGlobal@BT beats them.

Trials on *arXiv* Data. As an independent data set, we use our *arXiv* grahps for testing our results from $\mathcal{G}_e$ and the random instances. These graphs consist solely of glued cliques of authors (papers), established within single timesteps where potentially many new nodes and edges are introduced. Together with modularity's *resolution limit* [21] and its fondness of balanced clusters and a non-arbitrary number thereof in large graphs [29], these degenerate dynamics are adequate for fooling local algorithms that cannot regroup cliques all over as to modularity's liking: Static algorithms constantly reassess a growing component (Figs. 3a-3c), while dynamics using N or BN will sometimes have no choice but to further enlarge some growing cluster. Locally this is a good choice, but globally some far-away cut might qualify as an improvement over pure componentwise growth.

However, we measured that dGlobal@BT easily keeps up with the static algorithms' *modularity*, being able to adapt its number of clusters appropriately. The dynamic algorithms using other *prep strategies* do struggle to make up for their inability to re-cluster; however, they still only lag behind by about 1%. Figures 7 and 8 show *modularity* for coarse and fine batches, respectively, using the *arXiv* category *Nuclear Theory* (1992-2010, 33K e-prints, 200K elementary events, 14K authors). As before, dynamics are faster and smoother. For the coarse batches, speedups of 10 to 2K (BT) are attained; for fine batches, these are 100 to 2K. In line with the above observations, their clusterings are slightly coarser (except for dGlobal@BT). See Apps. E.2 and E.3 for insightful further results that exhibit dGlobal@BT's appropriateness.

Summary of Insights. The outcomes of our evaluation are very favorable for the dynamic approach in terms of all three criteria. Furthermore, the dynamics exhibit the ability to react quickly and adequately (Fig. 9) to changes in the random generator's ground-truth clustering.

We observed that dLocal is less susceptible to an increase of the search space than dGlobal. However, our results argue strongly for the locality assumption in both cases—an increase in the search space beyond a very limited range is not justified when trading off runtime against quality. On the contrary, quality and smoothness may even suffer for dLocal. Consequently, N and BN strategies with a limited range are capable of producing high-quality clusterings while excelling at smoothness. The BT strategy for dGlobal yields competitive quality at unrivaled speed, but at the expense of smoothness.

For dLocal a gradual improvement of quality and smoothness over time is observable, which can be interpreted as an effect reminiscent of *simulated annealing*, a technique that has been shown to work well for *modularity* maximization [28]. Our data indicates that the best choice for an algorithm in terms of quality may also depend on the nature of the target graph. While dLocal surpasses dGlobal on almost all generated graphs, dGlobal is superior on our real-world instance $\mathcal{G}_e$. We speculate that this is due to $\mathcal{G}_e$ featuring a power law degree distribution in contrast to the Erdős-Rényi-type generated instances. In turn, our *arXiv* trial graphs, which grow and shrink in a volatile but local manner, allow a for a small margin of quality improvement, if the clustering is regularly adapted globally (re-balanced and coarsened/refined). Only the statics and dGlobal@BT are able to do this, however, at the cost of smoothness. Universally, the latter algorithm is the fastest.

Concluding, some dynamic algorithm always beats the static algorithms; backtracking is preferable for locally concentrated or monotonic graph dynamics and a small search space is to be used for randomly distributed changes in a graph.

4 Conclusion

As the first work on *modularity*-driven clustering of dynamic graphs, we deal with the NP-hard problem of updating a *modularity*-optimal clustering after a change in the graph. We developed dynamizations of the currently fastest and the most widespread heuristics for *mod-*

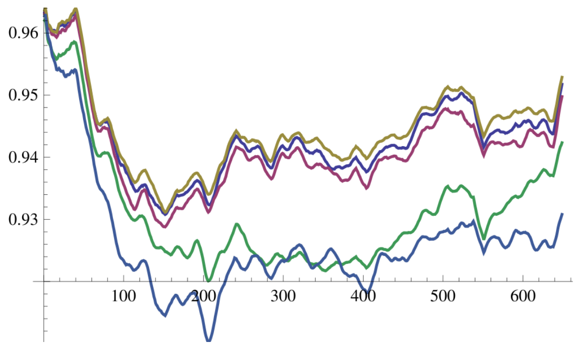

Fig. 7: *Modularity*, {arx}, batch size 50 e-prints: Backtracking (**dGlobal@BT**) easily follows the static algorithms (**sLocal** and **sGlobal**); even **dLocal@BN**$_4$ and **dGlobal@BN**$_{16}$ lag behind by only $\sim 1\%$.

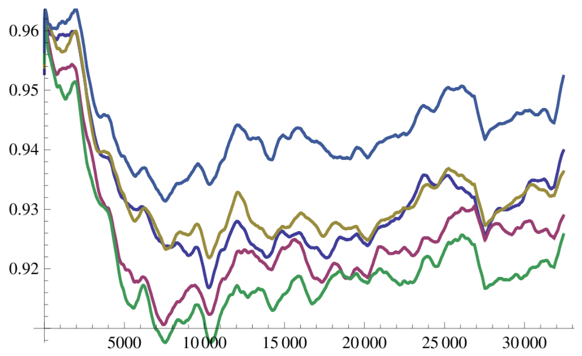

Fig. 8: *Modularity*, {arx}, batch size 1 e-print, dynamics only: **dGlobal@BT** excels, followed by **dLocal@N**$_1$ and **dLocal@BN**$_4$ and then **dGlobal@BN**$_{16}$ and **dGlobal@N**$_1$ which don't benefit from finer batches.

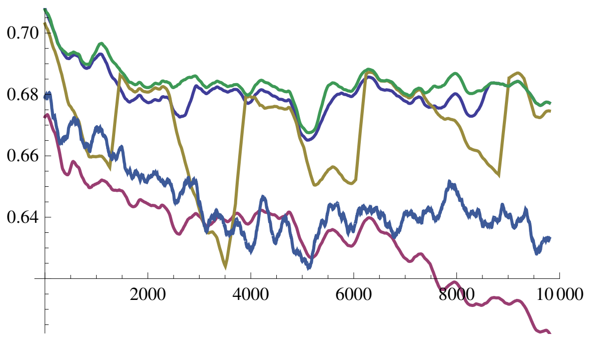

Fig. 9: *Modularity* on a long (10k time steps) slowly growing ($n = 1$k to 2k) random graph: The very simple method of the reference clustering to determine the ground-truth clustering relies solely on how clearly the split of a cluster, or the merge of two clusters has manifested in terms of the affected edge mass. Thus, the reference (yellow) exhibits jumps at such points, before which it clearly decreases. In contrast, the dynamic algorithms manage to adapt quickly to changes in the ground-truth clustering (from top to bottom: **dLocal@N**$_1$, **dLocal@BN**$_4$, **dGlobal@BT**, **dGlobal@BN**$_{16}$).

ularity-maximization and evaluated them and a dynamic partial ILP for local optimality. For our fastest update strategy, we can prove a tight bound of $\Theta(\log n)$ on the expected number of backtrack steps required. Our experimental evaluation on real-world dynamic networks and on dynamic clustered random graphs revealed that dynamically maintaining a clustering of a changing graph does not only save time, but also yields higher *modularity* than recomputation—except for degenerate graph dynamics—and guarantees much smoother clustering dynamics. Moreover, heuristics are better than being locally optimal at this task. Surprisingly small search spaces work best, avoid trapping local optima well and adapt quickly and aptly to changes in the ground-truth clustering, which strongly argues for the assumption that changes in the graph ask for local updates on the clustering.

References

1. Brandes, U., Delling, D., Gaertler, M., Görke, R., Höfer, M., Nikoloski, Z., Wagner, D.: On Modularity Clustering. IEEE Transactions on Knowledge and Data Engineering **20**(2) (February 2008) 172–188
2. Fortunato, S.: Community detection in graphs. Elsevier Physics Reports **486**(3–5) (2009)
3. Newman, M.E.J., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E **69**(026113) (2004)
4. Keogh, E., Lonardi, S., Ratanamahatana, C.A.: Towards Parameter-Free Data Mining. In: Proceedings of the 10th ACM SIGKDD International Conference, ACM Press (2004) 206–215
5. Schaeffer, S.E., Marinoni, S., Särelä, M., Nikander, P.: Dynamic Local Clustering for Hierarchical Ad Hoc Networks. In: Proc. of Sensor and Ad Hoc Communications and Networks, 2006. Volume 2., IEEE 667–672
6. Blondel, V., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment **2008**(10)
7. Delling, D., Görke, R., Schulz, C., Wagner, D.: ORCA Reduction and ContrAction Graph Clustering. In Goldberg, A.V., Zhou, Y., eds.: Proc. of the 5th Int. Conf. on Algorithmic Aspects in Information and Management (AAIM'09). Volume 5564 of LNCS., Springer (June 2009) 152–165
8. Görke, R., Hartmann, T., Wagner, D.: Dynamic Graph Clustering Using Minimum-Cut Trees. In: Algorithms and Data Structures, 11th Int. Workshop. Volume 5664 of LNCS., Springer (August 2009)
9. Hopcroft, J.E., Khan, O., Kulis, B., Selman, B.: Tracking Evolving Communities in Large Linked Networks. Proceedings of the National Academy of Science of the United States of America **101** (April 2004)
10. Palla, G., Barabási, A.L., Vicsek, T.: Quantifying social group evolution. Nature **446** (April 2007) 664–667
11. Aggarwal, C.C., Yu, P.S.: Online Analysis of Community Evolution in Data Streams. [30]
12. Sun, J., Yu, P.S., Papadimitriou, S., Faloutsos, C.: GraphScope: Parameter-Free Mining of Large Time-Evolving Graphs. In: Proc. of the 13th ACM SIGKDD Int. Conference, ACM Press (2007) 687–696
13. Hübner, F.: The Dynamic Graph Clustering Problem - ILP-Based Approaches Balancing Optimality and the Mental Map. Master's thesis, Universität Karlsruhe (TH), Fakultät für Informatik (May 2008)
14. Chakrabarti, D., Kumar, R., Tomkins, A.S.: Evolutionary Clustering. In: Proceedings of the 12th ACM SIGKDD International Conference, ACM Press (2006) 554–560
15. Schaeffer, S.E.: Graph Clustering. Computer Science Review **1**(1) (August 2007) 27–64
16. White, S., Smyth, P.: A Spectral Clustering Approach to Finding Communities in Graphs. [30] 274–285
17. Pons, P., Latapy, M.: Computing Communities in Large Networks Using Random Walks. Journal of Graph Algorithms and Applications **10**(2) (2006) 191–218
18. van Dongen, S.M.: Graph Clustering by Flow Simulation. PhD thesis, University of Utrecht (2000)
19. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Physical Review E **70**(066111) (2004)
20. Brandes, U., Erlebach, T., eds.: Network Analysis: Methodological Foundations. Volume 3418 of Lecture Notes in Computer Science. Springer (February 2005)
21. Fortunato, S., Barthélemy, M.: Resolution limit in community detection. PNAS **104**(1) (2007) 36–41
22. Newman, M.E.J.: Analysis of Weighted Networks. Physical Review E **70**(056131) (2004) 1–9
23. Görke, R., Gaertler, M., Hübner, F., and Wagner, D.: Computational Aspects of Lucidity-Driven Graph Clustering. *Journal of Graph Algorithms and Applications*, **14**(2) (2010)
24. Delling, D., Gaertler, M., Görke, R., Wagner, D.: Engineering Comparators for Graph Clusterings. In: Proc. of the 4th International Conference on Algorithmic Aspects in Information and Management (AAIM'08). Volume 5034 of LBCS., Springer (June 2008) 131–142

25. Noack, A., Rotta, R.: Multi-level Algorithms for Modularity Clustering. In Vahrenhold, J., ed.: Proc. of the 8th Int. Symposium on Exp. Algorithms. Volume 5526 of LNCS., Springer (June 2009) 257–268
26. Görke, R., Staudt, C.: A Generator for Dynamic Clustered Random Graphs. Technical report, ITI Wagner, Faculty of Informatics, Universität Karlsruhe (TH) (2009) Informatik, TR 2009-7.
27. Brandes, U., Gaertler, M., Wagner, D.: Experiments on Graph Clustering Algorithms. In: Proc. of the 11th Annual European Symposium on Algorithms (ESA'03). Volume 2832 of LNCS., Springer (2003) 568–579
28. Guimerà, R., Amaral, L.A.N.: Functional Cartography of Complex Metabolic Networks. Nature **433** (February 2005) 895–900
29. Good, B. H., de Montjoye, Y. and Clauset, A.: The performance of modularity maximization in practical contexts. arxiv.org/abs/0910.0165 (2009)
30. Proceedings of the fifth SIAM International Conference on Data Mining. In: Proceedings of the fifth SIAM International Conference on Data Mining, SIAM (2005)

Appendix

A Implementation Notes.

We conducted our experiments on up to eight cores, 1 per experiment, of a dual *Intel Xeon* 5430 running *SUSE Linux* 11.1. The machine is clocked at 2.6GHz, has 32GB of RAM and $2\times$1MB of L2 cache. Our algorithms and measures are implemented in *Java* 1.6.0_13, partially using the *yFiles* graph library[9], and run on a 64-Bit Server VM. Evaluations, plots and the setups of experiments were conducted via a frontend programmed in *Mathematica* (version 7.0.1.0). As priority queue we use a `java.util.PriorityQueue`. As a data structure which supports *backtrack*, instead of using a rather involved fully dynamic *union-find* structure, we maintain a similar structure, a binary forest with actual nodes as leaves and the merge operations as internal tree-nodes.

B Full and Partial ILP Formulations

We realize a distance relation between elements—nodes to start with—of the graph, which indicates whether elements belong together in the clustering. We denote the set of *node distance* variables (see Eq. 2) by

$$\mathcal{X}(\tilde{V}) := \{X_{uv} : \{u,v\} \in \binom{\tilde{V}}{2}\} \quad \text{with} \quad X_{uv} = \begin{cases} 0 & \text{if } \mathcal{C}(u) = \mathcal{C}(v) \\ 1 & \text{otherwise} \end{cases} . \tag{5}$$

With these a full ILP formulation of *modularity*-optimization is already possible if we set $\tilde{V} = V$, i.e., all nodes are "free". We merely have to ensure the properties of an equivalence relation, *reflexivity*, *symmetry* and *transitivity*.

$$\forall \{u,v,w\} \in \binom{\tilde{V}}{3} : \begin{cases} X_{uv} + X_{vw} - X_{uw} \geq 0 \\ X_{uv} + X_{uw} - X_{vw} \geq 0 \\ X_{uw} + X_{vw} - X_{uv} \geq 0 \end{cases} \quad ; \quad X_{uv} \in \{0,1\} \tag{6}$$

$$\text{minimize } \mathrm{mod}_{\mathrm{ILP}}(G, \mathcal{C}_G) = \sum_{\{u,v\} \in \binom{V}{2}} \left(\omega(u,v) - \frac{\omega(u) \cdot \omega(v)}{2 \cdot \omega(E)} \right) X_{uv} \tag{7}$$

Equation 6 represents transitivity, we can omit the other two: Reflexivity, $X_{uu} = 0$, is automatically ensured since a node is always in the same cluster as itself. Symmetry, $X_{uv} = X_{vu}$, is ensured since there is only one such variable.

Elements are Nodes and Preserved Clusters. For the partial ILP we usually preserve some clusters $\tilde{\mathcal{C}}$ and have only some free nodes, so $\tilde{V} \subseteq V$. Nodes are allowed to join other clusters and to form new ones, but preserved clusters can neither split nor merge. To indicate whether a free node joins a cluster we introduce the set of *node-cluster distance* variables

$$\mathcal{Y}(\tilde{V}, \tilde{\mathcal{C}}) := \{Y_{uC} : \{u, C\} \in \tilde{V} \times \tilde{\mathcal{C}}\} \quad \text{with} \quad Y_{uC} = \begin{cases} 0 & \text{if } \mathcal{C}(u) = C \\ 1 & \text{otherwise} \end{cases} . \tag{8}$$

[9]Licensed from *yWorks* for more information, see www.yworks.com .

We now need to couple these variables with $\mathcal{X}$ to ensure that if two nodes u, v join the same cluster, their variable X_{uv} also reflects that they are clustered together. Moreover a node must only join one cluster, and the target function must evaluate such joins:

$$\forall \{u,v,w\} \in \binom{\tilde{V}}{2} \times \tilde{\mathcal{C}} : \begin{cases} X_{uv} + Y_{uC} - Y_{vC} \geq 0 \\ X_{uv} + Y_{vC} - Y_{uC} \geq 0 \\ Y_{uC} + Y_{vC} - X_{uv} \geq 0 \end{cases} \quad ; \quad Y_{uC} \in \{0,1\} \tag{9}$$

$$\forall u \in \tilde{V} : \sum_{C \in \tilde{\mathcal{C}}} Y_{uC} \geq k-1 \quad \text{(a node's cluster must be unique)} \tag{10}$$

$$\begin{aligned} \text{minimize } \operatorname{mod}_{\substack{\mathsf{partialILP} \\ \text{no merge}}}(G, \mathcal{C}) = & \sum_{X_{uv} \in \mathcal{X}(\tilde{V})} \left(\omega(u,v) - \frac{\omega(u) \cdot \omega(v)}{2\omega(E)} \right) X_{uv} \\ & + \sum_{Y_{uC} \in \mathcal{Y}(\tilde{V}, \tilde{\mathcal{C}})} \left(\sum_{w \in C} \left(\omega(u,w) - \frac{\omega(u) \cdot \omega(w)}{2\omega(E)} \right) \right) Y_{uC} \end{aligned} \tag{11}$$

Preserved Clusters may Merge. Finally, if we also allow pre-clusters to merge, we can handle them just as we handle nodes. We thus additionally introduce *cluster distance* variables, which indicate whether two clusters merge:

$$\mathcal{Z}(\mathcal{C}) := \{ Z_{CD} : \{C, D\} \in \binom{\tilde{\mathcal{C}}}{2} \} \quad \text{with} \quad Z_{CD} = \begin{cases} 0 & \mathsf{merge}(C,D) \\ 1 & - \end{cases} \tag{12}$$

To ensure consistency we need constraints as in Eq. 3 for $\mathcal{Z}$. Additionally, just as for $\mathcal{X}$ we need to couple $\mathcal{Z}$ with $\mathcal{Y}$, and let the target function evaluate merging clusters. In turn we must now drop the constraints in Eq. 10, since now a node *can* join more than one cluster—if and only if these clusters merge.

$$\forall \{C, D, E\} \in \binom{\tilde{\mathcal{C}}}{3} : \begin{cases} Z_{CD} + Z_{DE} - Z_{CE} \geq 0 \\ Z_{CD} + Z_{CE} - Z_{DE} \geq 0 \\ Z_{CE} + Z_{DE} - Z_{CD} \geq 0 \end{cases} \tag{13}$$

$$\forall \{u, C, D\} \in \tilde{V} \times \binom{\tilde{\mathcal{C}}}{2} : \begin{cases} Z_{CD} + Y_{uD} - Y_{uC} \geq 0 \\ Z_{CD} + Y_{uC} - Y_{uD} \geq 0 \\ Y_{uC} + Y_{uD} - Z_{CD} \geq 0 \end{cases} \tag{14}$$

$$\begin{aligned} \text{minimize } \operatorname{mod}_{\substack{\mathsf{partialILP} \\ \text{merge}}}(G, \mathcal{C}) = & \sum_{X_{uv} \in \mathcal{X}(\tilde{V})} \left(\omega(u,v) - \frac{\omega(u) \cdot \omega(v)}{2\omega(E)} \right) X_{uv} \\ & + \sum_{Y_{uC} \in \mathcal{Y}(\tilde{V}, \tilde{\mathcal{C}})} \left(\sum_{w \in C} \left(\omega(u,w) - \frac{\omega(u) \cdot \omega(w)}{2\omega(E)} \right) \right) Y_{uC} \\ & + \sum_{Z_{CD} \in \mathcal{Z}(\tilde{\mathcal{C}})} \left(\sum_{x \in C} \sum_{y \in D} \left(\omega(x,y) - \frac{\omega(x) \cdot \omega(y)}{2\omega(E)} \right) \right) Z_{CD} \end{aligned} \tag{15}$$

Summary. In Table 5 we summarize which constraints are necessary for which problem formulation. Preliminary experiments using techniques such as breaking symmetry, orbitopal fixing or lazy constraints did not seem promising although a thorough investigation might yield some mild speedup. Note that for the case where merging is allowed we could also have variables as in Eq. 6 for $\mathcal{Z} \cup \mathcal{X}$, and discard $\mathcal{Y}$ altogether. Note further that if in addition to the merging of clusters we also allow splitting, we actually arrive at the full ILP again.

Table 5: ILP variants

Name	Constraint Set
Full ($\tilde{V} = V$)	5-6
noMerge	5-6, 8-10
merge	5-6, 8-9, 12-14

C The Backtrack Strategy

To motivate the *backtack strategy* we first detail some insights on the change in *modularity* if (i) the graph changes and (ii) we decide to move some nodes from one cluster to another, in order to react to the change. Please note that all statements generalize trivially to weighted edges. Let $C \in \mathcal{C}$ be a cluster and $D \in \{\mathcal{C} \cup \emptyset\}$ be a cluster or the empty set. Let further $U \subset C$ be a subset of C, and define further the clustering $\mathcal{D}$:

$$\mathcal{D} := (\mathcal{C} \setminus \{C, D\}) \cup \{C \setminus U, D \cup U\} \quad (\text{move } U \text{ from } C \text{ to } D) \tag{16}$$

The basis of *modularity* (*Mod*), *coverage* (*Cov*) and the expected value of *coverage* (*ECov*) change when we move from clustering $\mathcal{C}$ to $\mathcal{D}$; we can express these changes and the change Δ in *modularity* as follows:

$$\Delta_{Cov} := Cov(\mathcal{D}) - Cov(\mathcal{C}) \ , \quad \Delta_{ECov} := ECov(\mathcal{D}) - ECov(\mathcal{C}) \ , \tag{17}$$

$$\Delta := Mod(\mathcal{D}) - Mod(\mathcal{C}) = \Delta_{Cov} - \Delta_{ECov} \tag{18}$$

Note that Δ must be non-positive if $\mathcal{C}$ was optimal. If we generalize the definitions in Sec. 1 from clusters to general sets of nodes, then we can write these as:

$$\Delta_{Cov} := \frac{E(U, D) - E(U, C \setminus U)}{m} \ , \tag{19}$$

$$\Delta_{ECov} := \frac{1}{4m^2} \left(\sum_{B \in \mathcal{D}} \deg^2(B) - \sum_{B \in \mathcal{C}} \deg^2(B) \right) \tag{20}$$

$$= \frac{1}{4m^2} \left(\deg^2(D \cup U) + \deg^2(C \setminus U) - \deg^2(D) - \deg^2(C) \right) \tag{21}$$

$$= \frac{\deg(U)}{2m^2} (\deg(D) - \deg(C \setminus U)) \tag{22}$$

Given a change in the graph we want to know whether moving from $\mathcal{C}$ to $\mathcal{D}$ is beneficial. Thus, in addition to moving from $\mathcal{C}$ to $\mathcal{D}$, we now move from G to G', i.e., we change graph G by, say, adding edge $\{v, w\}$. Analogously to the above we now define Δ'_{Cov}, Δ'_{ECov} and Δ'_{Mod}. We can now establish sufficient and necessary conditions for Δ' to be positive if $\Delta \leq 0$, the following two Tabs. 6-7. We distinguish cases whether or not v and w are elements of C, D or U. In both tables, this is done in the first column and the second columns give the appropriate values of Δ'_{Cov} and Δ'_{ECov}. The last columns give tight conditions for Δ' to be strictly positive, i.e., for the case when moving such a set U from C to D increases *modularity* in G'.

Table 6: The different effects on *modularity* if, after the *insertion* of edge $\{v, w\}$, we move a specific subset U of nodes from cluster C to cluster D.

preconditions	formulae for Δ'_{Cov} and Δ'_{ECov}	$\Delta \leq 0$ and $\Delta' > 0$ iff
$v, w \notin C \cup D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov})$
$v \in C \backslash U,\ w \notin D$ or $w \in C \backslash U,\ v \notin D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} - \frac{\deg_G(U)}{2(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} + \frac{\deg_G(U)}{2m^2})$
$v \in C \backslash U,\ w \in D$ or $w \in C \backslash U,\ v \in D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov})$
$v \notin C,\ w \in D$ or $w \notin C,\ v \in D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} + \frac{\deg_G(U)}{2(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} - \frac{\deg_G(U))}{2m^2})$
$v \in U,\ w \notin D$ or $w \in U,\ v \notin D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov})$
$v \in U,\ w \in D$ or $w \in U,\ v \in D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov} + \frac{1}{m+1}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} + \frac{\deg_G(U)}{2(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} + \frac{2m+2-\deg_G(U)}{2m^2})$

Table 7: For the *deletion* of edge $\{v, w\}$, this table details effects analogously to Tab. 6

preconditions	formulae for Δ'_{Cov} and Δ'_{ECov}	$\Delta \leq 0$ and $\Delta' > 0$ iff
$v, w \notin C \cup D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov})$
$v, w \in C \backslash U$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} - \frac{\deg_G(U)}{(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} + \frac{\deg_G(U)}{m^2})$
$v, w \in D$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} + \frac{\deg_G(U)}{(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} - \frac{\deg_G(U)}{m^2})$
$v, w \in U$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} + \frac{\deg_G(D)-\deg_G(C\backslash U)}{(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} - \frac{\deg_G(D)-\deg_G(C\backslash U)}{m^2})$
$v \in U,\ w \in C \backslash U$ or $w \in U,\ v \in C \backslash U$	$\Delta'_{Cov} = \frac{m}{m+1}\Delta_{Cov} - \frac{1}{m+1}$ $\Delta'_{ECov} = (\frac{m}{m+1})^2\Delta_{ECov} + \frac{\deg_G(D)-\deg_G(C)-1}{2(m+1)^2}$	$\Delta_{ECov} \in$ $[\Delta_{Cov}, (1+\frac{1}{m})\Delta_{Cov} - \frac{2m+1+\deg_G(D)-\deg_G(C)}{2m^2})$

Summarizing, if we want to adapt a clustering to a change in a graph by moving a set U of nodes between clusters, the given ranges for $\mathit{\Delta}_{ECov}$ categorize exactly when a given set U (specified by the first column) will increase *modularity* for the new graph G'. However, this does not determine a *specific* set U—we still have to decide on this, but by the size of the range for $\mathit{\Delta}_{ECov}$ we can deduce some structure. Since we aim at a dynamization dGlobal of the global agglomerative algorithm, a reasonable approach is as follows: Backtrack specific merge operations of the static algorithm sGlobal until the most promising (according to Tabs. 6-7) operations in terms of moving a set U are available; then let the algorithm finish the clustering for G'. Of course this does not yield optimality by any means, nor does it *identify* the best set U, but it gives Global a fighting chance to find a good improvement with minimum effort, since exclusively the most promising parts of the clustering are broken up.

In the following we detail our update procedures which are motivated by the above discussion. For these we require the helper algorithms given in Algs. 3-5. Algorithm 3, backtrack(v), splits the cluster containing v into those two parts it resulted from.

```
Algorithm 3: backtrack(v)
1 $\mathcal{C}(v) \rightarrow \mathcal{C}'(v) \cup \mathcal{C}(v) \setminus \mathcal{C}'(v)$
```

Algorithm 4, isolate(v), iteratively backtracks those merges that involved v, until v is contained in a singleton. Algorithm 5, separate(u, v), backtracks those merges involving both u and v until u and v are in different clusters.

Algorithm 4: isolate(v)

```
1 while $|\mathcal{C}(v)| \neq 1$ do
2     backtrack($v$)
```

Algorithm 5: separate(u, v)

```
1 while $\mathcal{C}(u) = \mathcal{C}(v)$ do
2     backtrack($u$)
```

Backtrack Inter-Cluster Edge Addition. Since we assume for most graphs that the degree of each cluster does not exceed m, we have the best chances to increase *modularity* if we choose U to contain either v or w and move U to the cluster containing the other node (see sixth case in Tab. 6). Therefore, we define this part of our backtrack strategy as isolate(u), isolate(v).

Backtrack Intra-Cluster Edge Addition. In this case it seems to be the best choice to set U in such a way that it is a subset of C and contains either one of v or w, or both (cases two and four in Tab. 6; regarding case four, note that D can be empty, which implies $\deg_G(D) = 0$). We thus define this backtrack case as separate(u, v). Since, however, the expected number of backtrack operations is 2 if we assume that in each such split, u and v are separated with probability 1/2 (see below for details), one might think that this is too little an invested effort. We thus also tested an alternative which performs isolate(u) and isolate(v); however this uniformly did not raise quality but only runtime and distance.

Backtrack Edge Deletions. For the case of edge deletions it is more difficult to find good backtracking strategies. For additions we can try to reasonably reduce the sizes of the affected clusters by splitting them into parts that either form new clusters or merge with existing ones. The strategy in the case of deletions should thus be the inverse: Split off parts of the clusters that are unaffected by the edge deletion and link them with the affected ones. But how do we know which cluster to split? We leave this question unanswered, analytically, and rely on common sense to define the following procedures:

- Inter-cluster edge deletion: *do nothing* ($\tilde{\mathcal{C}} = \mathcal{C}$, but do call Global, as usual)
- Intra-cluster edge deletion: isolate(u), isolate(v)

Analysis of the isolate and the separate Operations. It is easy to see that the expected number $E\{S\}$ of backtrack steps S for a single call of separate(u, v) is 2, if we assume that a backtrack step divides a cluster randomly and thus separates u and v with probability 1/2. Without further a priori knowledge this is a reasonable assumption; however, it is crucial to note that all our findings (The. 1 in particular) remain valid for any arbitrary but fixed constant probability instead of 1/2. For simplicity we use 1/2 in the following. Then, S is distributed according to the geometric distribution with parameter 1/2 yielding $E\{S\} = 2$.

For the proof that the expected number $E\{I\}$ of backtrack steps I for a single call of isolate(v) is in $\Theta(\ln n)$ (see The. 1), we require the following two lemmas for the theorem that proves the bound.

Lemma 1. *Let $(\Omega, \mathcal{A}, P)$ be a probability space, $A_1, \ldots, A_n \in \mathcal{A}$ independent events with $P(A_i) = p, i = 1, \ldots, n$. Then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 1 - (1-p)^n \quad .$$

(Proof omitted)

Lemma 2. *Let $i \in \mathbb{N}_0, j \in \mathbb{N}$. Then it holds that*

$$\int_0^\infty \left(\frac{1}{2}\right)^{jx} \left(1 - \left(\frac{1}{2}\right)^x\right)^i dx = \frac{i!(j-1)!}{(j+i-1)!} \cdot \frac{1}{\ln 2} \cdot \frac{1}{i+j} \quad .$$

Specifically, for $j = 1$ the following equation holds:

$$\int_0^\infty \left(\frac{1}{2}\right)^{x} \left(1 - \left(\frac{1}{2}\right)^x\right)^i dx = \frac{1}{\ln 2} \cdot \frac{1}{i+1}$$

Proof. The proof uses induction over i, with partial integration for the induction step. Note that it holds for all $j \in \mathbb{N}$. For brevity we just give a proof sketch.

$$\begin{aligned}
&\int_0^\infty \underbrace{\left(\frac{1}{2}\right)^{jx}}_{g'} \underbrace{\left(1 - \left(\frac{1}{2}\right)^x\right)^i}_{f} dx \\
&\text{(such that } g = -\frac{1}{j \ln 2}\left(\frac{1}{2}\right)^{jx} \text{ and } f' = \ln 2 \left(\frac{1}{2}\right)^x i \left(1 - \left(\frac{1}{2}\right)^x\right)^{i-1} \quad) \\
&= \left[-\frac{1}{j \ln 2}\left(\frac{1}{2}\right)^{jx} \cdot \left(1 - \left(\frac{1}{2}\right)^x\right)^i\right]_0^\infty \quad (= 0 \text{ for } i \neq 0, \text{ which holds here}) \\
&- \int_0^\infty -\frac{1}{j \ln 2}\left(\frac{1}{2}\right)^{jx} \cdot \ln 2 \left(\frac{1}{2}\right)^x i \left(1 - \left(\frac{1}{2}\right)^x\right)^{i-1} dx \\
&= 0 + \frac{i}{j}\int_0^\infty \left(\frac{1}{2}\right)^{(j+1)x} \left(1 - \left(\frac{1}{2}\right)^x\right)^{i-1} dx \\
&\quad \vdots \\
&\frac{i \cdot \ldots \cdot 1}{j \cdot \ldots \cdot j+i-1} \cdot \int_0^\infty \left(\frac{1}{2}\right)^{(i+j)x} dx \\
&\frac{i!(j-1)!}{(j+i-1)!} \cdot \frac{1}{\ln 2} \cdot \frac{1}{i+j}
\end{aligned}$$

Theorem 2. *Let $n \in \mathbb{N}$ and $X_j^{(i)}, i = 1, \ldots, n,\ j = 1, 2, \ldots$ be i.i.d. random variables that are Bernoulli-distributed with parameter $\frac{1}{2}$. We define*

$$N := \min\{k \in \mathbb{N}_0 : \forall i \in \{2, \ldots, n\}\ \exists j \in \{1, \ldots, k\} : X_j^{(i)} \neq X_j^{(1)}\} \quad .$$

Then it follows for the expectance of N:

$$E\{N\} \in \Theta(\ln n)$$

Proof. W.l.o.g. $n \geq 2$.

$$\begin{aligned}
E\{N\} &= \sum_{k=0}^{\infty} P(N = k) \cdot k = \sum_{k=0}^{\infty} P(N > k) \\
&= \sum_{k=0}^{\infty} P(\exists i \in \{2, \ldots, n\}\ \forall j \in \{1, \ldots, k\} : X_j^{(i)} = X_j^{(1)}) \\
&\text{(set event } A_i : \forall j \in \{1, \ldots, k\} : X_j^{(i)} = X_j^{(1)}\text{, then it holds } P(A_i) = \left(\frac{1}{2}\right)^k\text{)} \\
&= \sum_{k=0}^{\infty} 1 - \left(1 - \left(\frac{1}{2}\right)^k\right)^{n-1} && \text{(Lemma 1)} \\
&= \sum_{k=0}^{\infty} \frac{1 - \left(1 - (\frac{1}{2})^k\right)^{n-1}}{1 - \left(1 - (\frac{1}{2})^k\right)} \left(\frac{1}{2}\right)^k && \text{(rewrite)} \\
&= \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \sum_{i=0}^{n-2} \left(1 - \left(\frac{1}{2}\right)^k\right)^i && \text{(geom. series)} \\
&= \sum_{i=0}^{n-2} \sum_{k=0}^{\infty} \left(\frac{1}{2}\right)^k \cdot \left(1 - \left(\frac{1}{2}\right)^k\right)^i && \text{(reorder)} \\
&\in \Theta\left(\sum_{i=0}^{n-2} \int_0^{\infty} \left(\frac{1}{2}\right)^x \cdot \left(1 - \left(\frac{1}{2}\right)^x\right)^i dx\right) && \text{(rect. approx.)} \\
&\in \Theta\left(\sum_{i=0}^{n-2} \frac{1}{\ln 2} \cdot \frac{1}{i+1}\right) && \text{(Lemma 2)} \\
&\in \Theta\left(\frac{1}{\ln 2} \sum_{i=0}^{n-2} \frac{1}{i+1}\right) \in \Theta\left(\frac{1}{\ln 2}\left(\sum_{i=1}^{n} \frac{1}{i} - \frac{1}{n}\right)\right) \\
&\in \Theta(\ln n) && \text{(n-th harmonic no.)}
\end{aligned}$$

We can interpret the random variables $X_j^{(i)} \in \{0, 1\}$ of experiments X_j such that for the j-th division (by a backtrack step), $X_j^{(i)} = 0$ if the i-th node is in the left half, and $X_j^{(i)} = 1$ if the i-th node is in the right half. If we assign to node v index 1, then $X_j^{(i)} = X_j^{(1)}$ means, that node i is in the same half as v. Event A_i then means that for all experiments $1, \ldots, k$ node i always ended up on the same half as v. We thus look for the first experiment such that each node other than v has ended up in another half than v at least once (note that multiply separating a node i from v doe not alter the statement). Now its easy to see that The. 2 proves The. 1.

D Further Experiments

D.1 Fundamental Observations

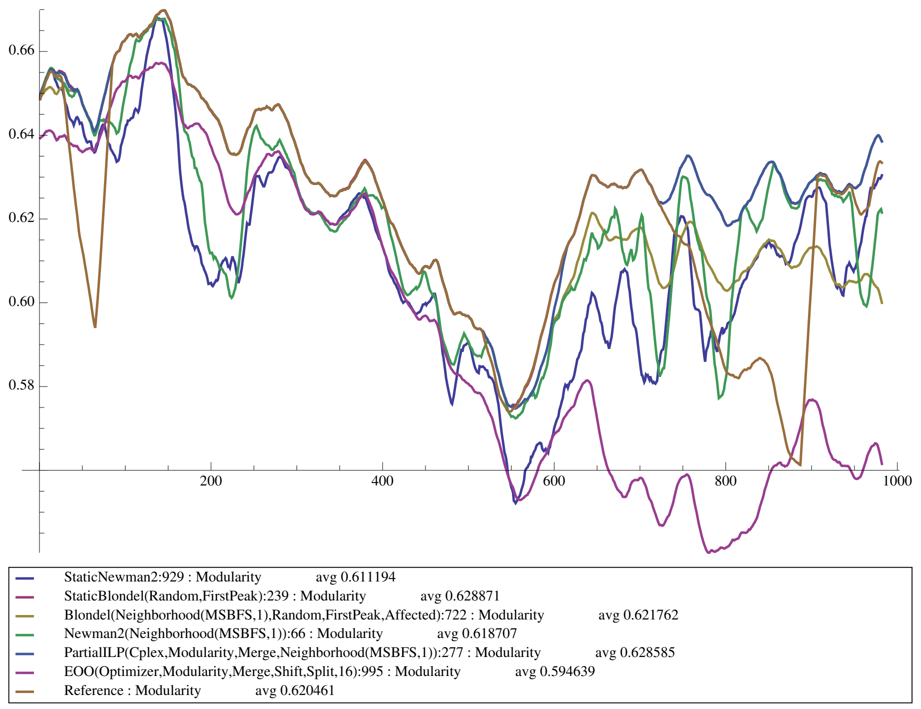

Fig. 10: In terms of *modularity*, **dEOO** lags behind the other algorithms.

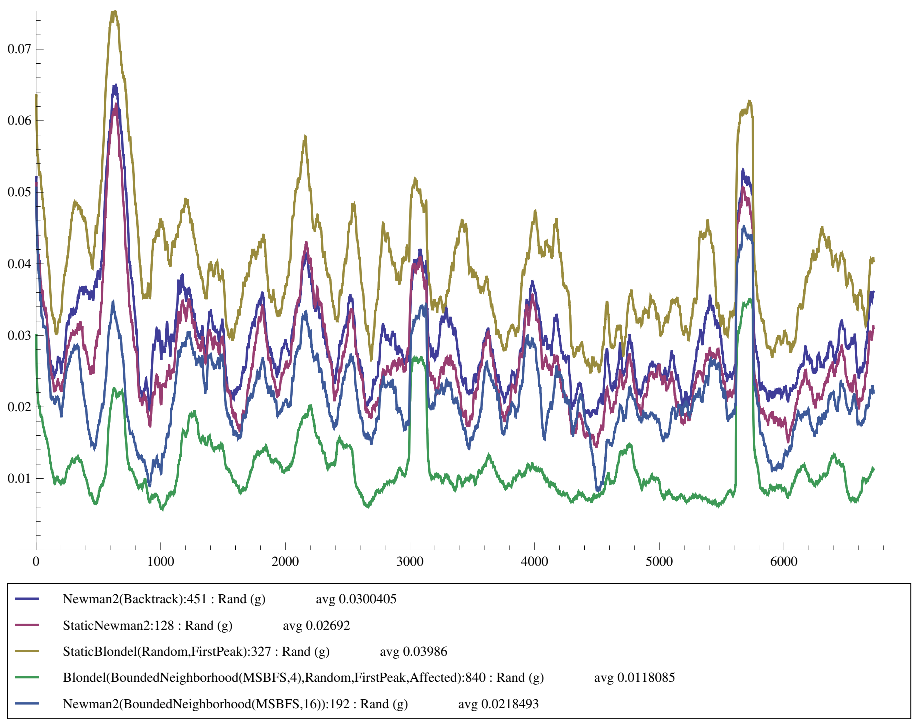

Fig. 11: In terms of $\mathcal{R}_g$-distance, dynamics are consistently better on $\mathcal{G}_e$.

D.2 Performance of sGlobal and sLocal

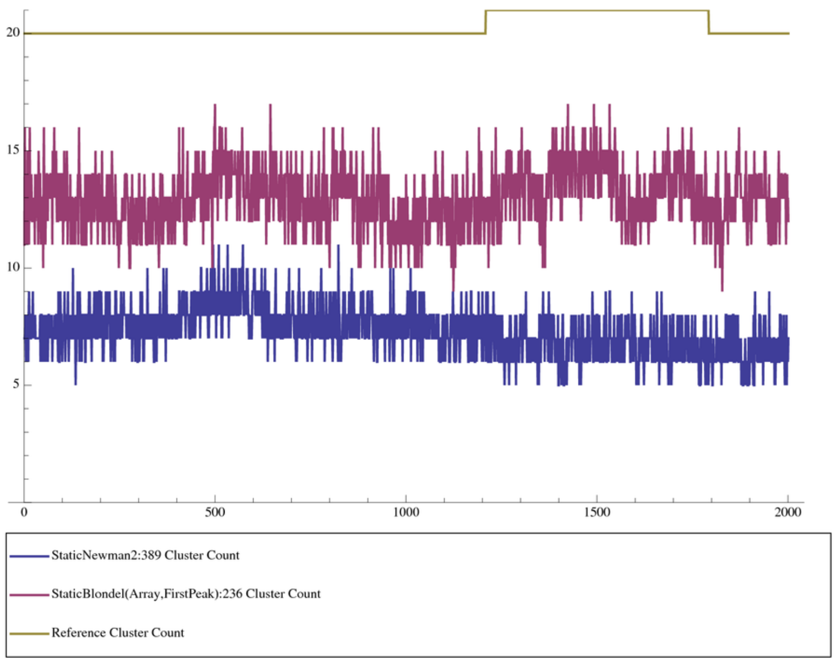

Fig. 12: Cluster count: sLocal yields a finer clustering than sGlobal, a similar observation holds for the dynamic counterparts.

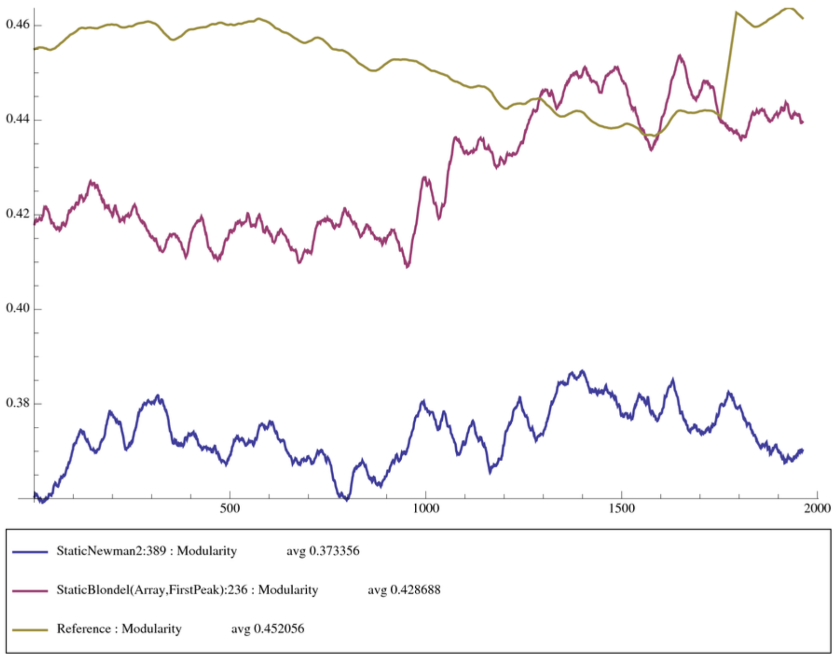

Fig. 13: Quality: sLocal surpasses sGlobal on this generated graph.

D.3 Node Orders of **dLocal** and **sLocal**

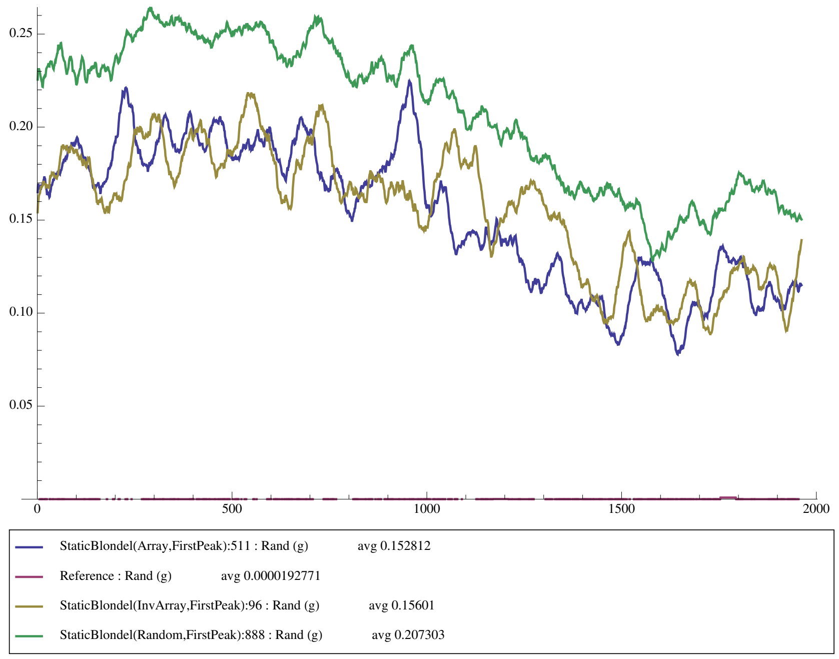

Fig. 14: The effect of different node orders, **Random** and two fixed orders (**Array** and **InvArray**), for **sLocal** on smoothness, in terms of $\mathcal{R}_g$-distance.

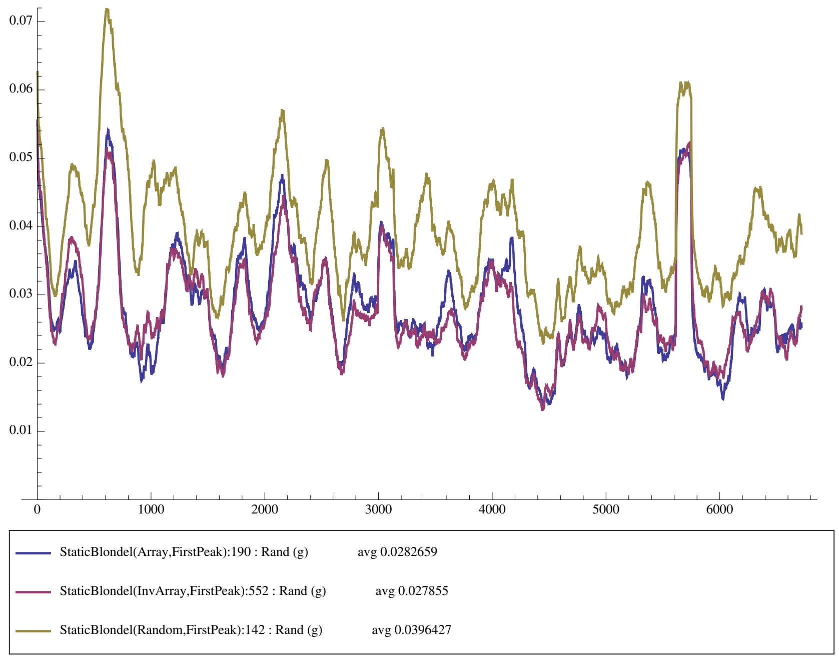

Fig. 15: The effect of different node orders, **Random** and two fixed orders (**Array** and **InvArray**), for **dLocal** on smoothness, in terms of $\mathcal{R}_g$-distance.

D.4 On the Behavior of the Dynamics

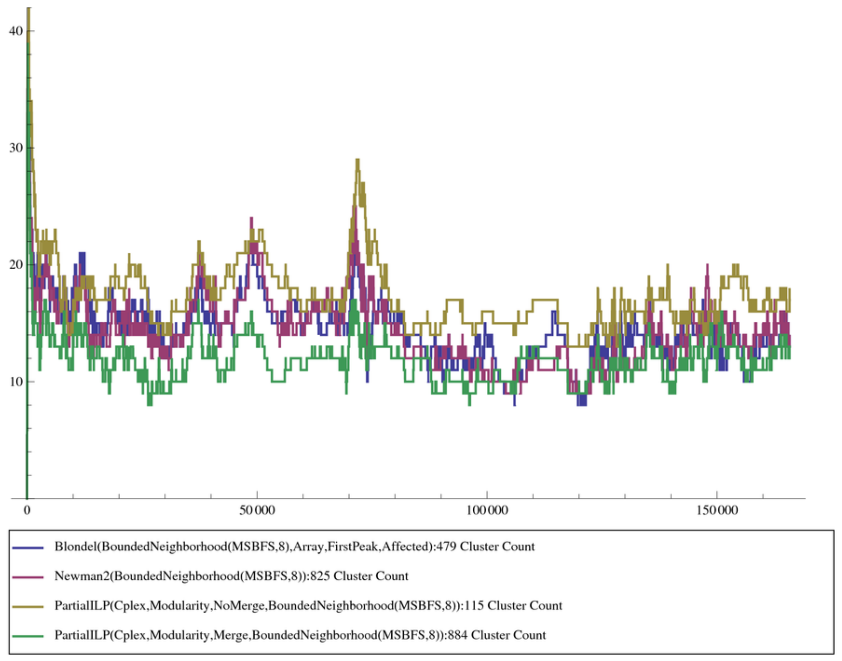

Fig. 16: Cluster count: **dILP(merge)** and **dILP(noMerge)** roughly bound **dLocal** and **dGlobal** from below and above, respectively

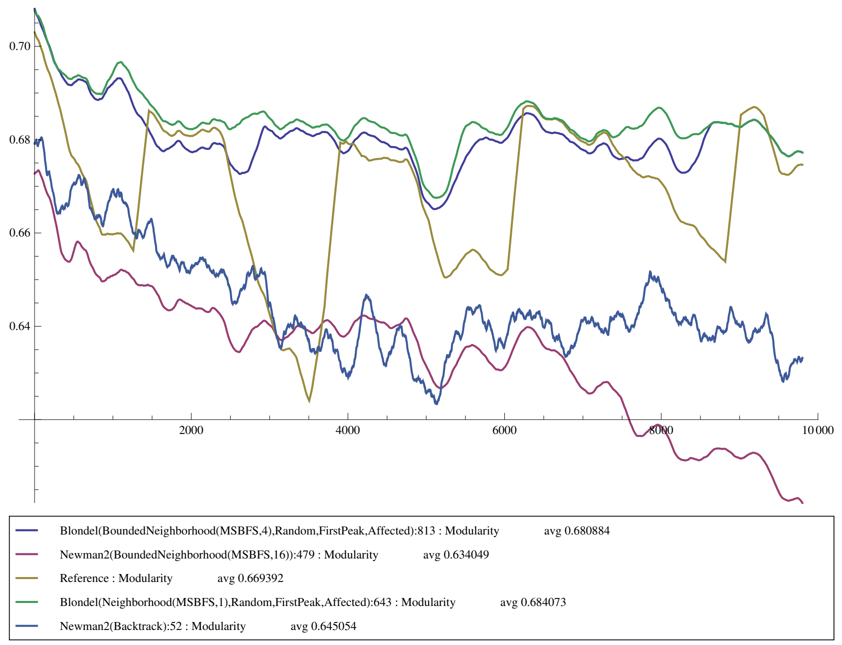

Fig. 17: Dynamic algorithms (from top to bottom: **dLocal@N**$_1$, **dLocal@BN**$_4$, **reference**, **dGlobal@BT**, **dGlobal@BN**$_{16}$) adapt quickly to changes in the ground-truth clustering, such changes are indicated by drops in the **reference** quality.

E Different Distance Measures Agree

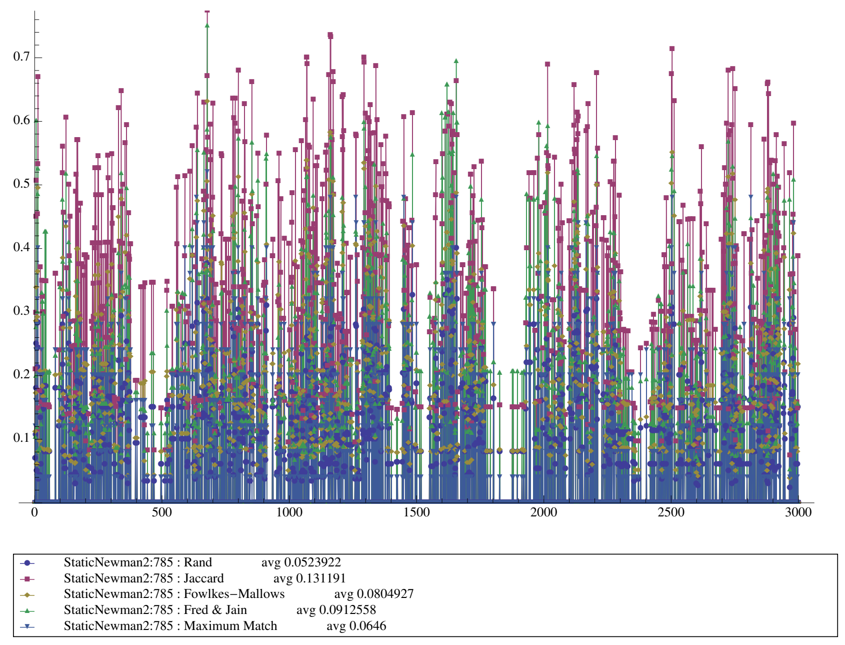

Fig. 18: Raw data for several distance measures

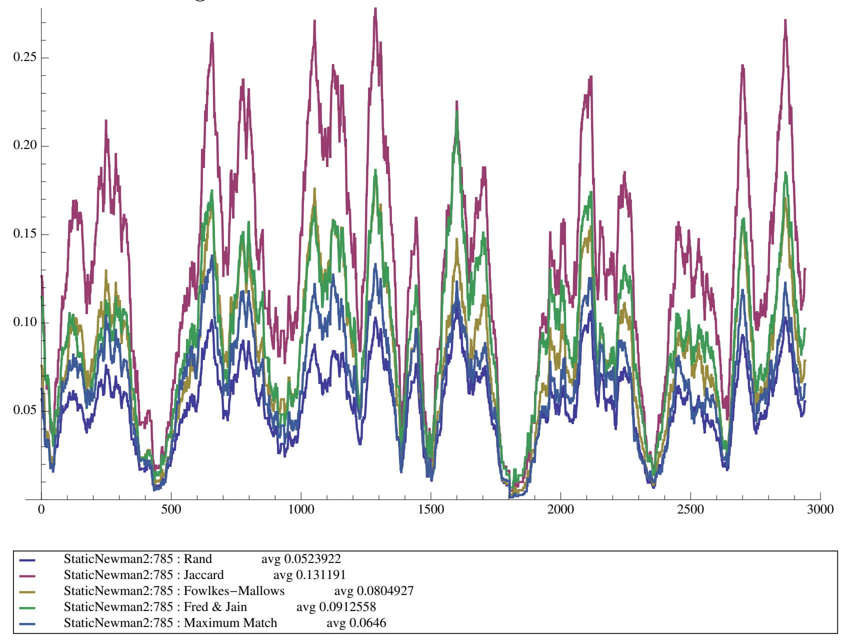

Fig. 19: Smoothed data corresponding to Fig. 18

E.1 Prep Strategies: Comparison of Sizes s for $\mathsf{dGlobal@BN}_s$

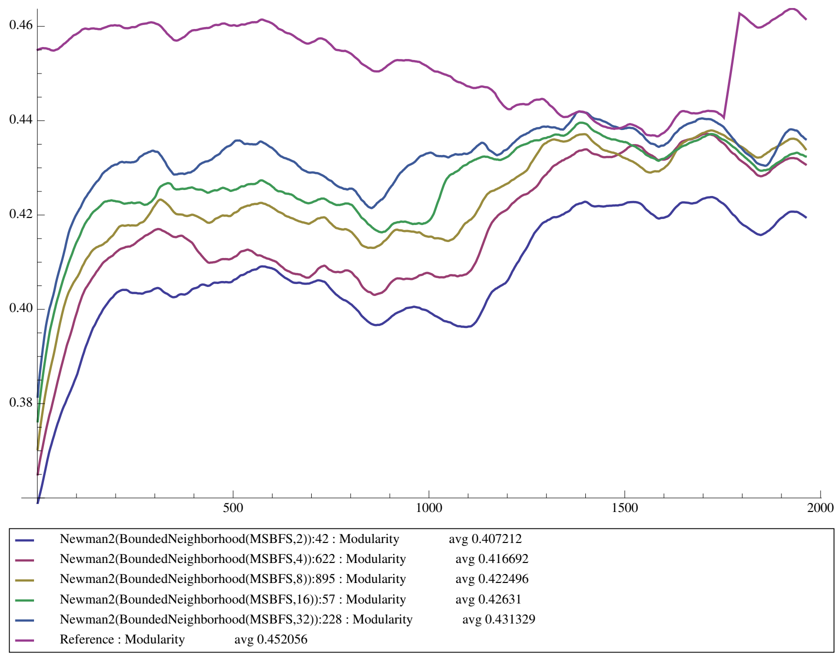

Fig. 20: Quality: $\mathsf{dGlobal@BN}_s$ benefits from higher values of s in this range

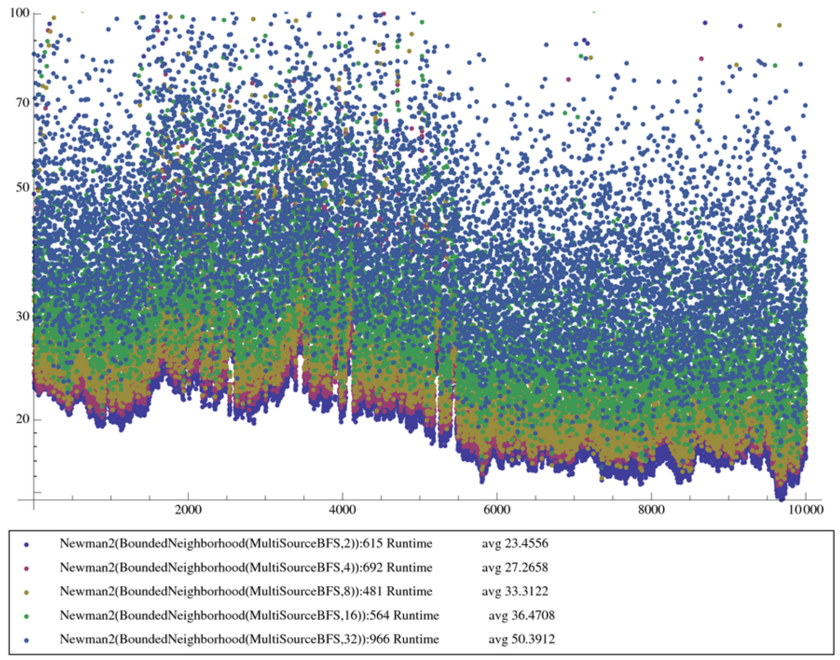

Fig. 21: Runtime for $\mathsf{dGlobal@BN}_s$ increases (sublinearly) with s

E.2 Trials with arXiv Data: Category *Nuclear Theory*

Dynamics and statics, *batch size* 50:

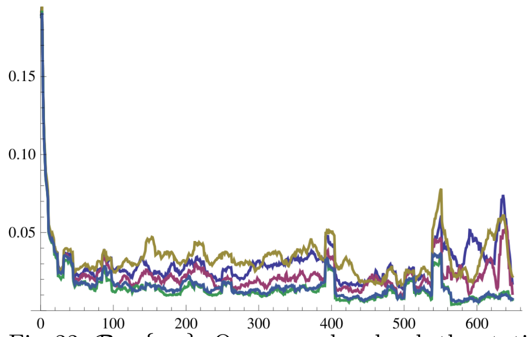

Fig. 22: $\mathcal{R}_g$, {arx}: On a very low level, the statics (**sLocal** and **sGlobal**) and **dGlobal@BT** are slightly less smooth than **dLocal@BN**$_4$ and **dGlobal@BN**$_{16}$. The overall level indicates the graph's stability.

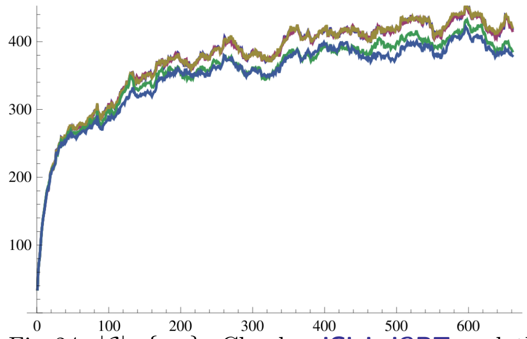

Fig. 24: $|\mathcal{C}|$, {arx}: Clearly, **dGlobal@BT** and the statics (**sLocal** and **sGlobal**) rebalance and reorganize the clustering, while **dLocal@BN**$_4$ and **dGlobal@BN**$_{16}$ more often enlarge existing clusters.

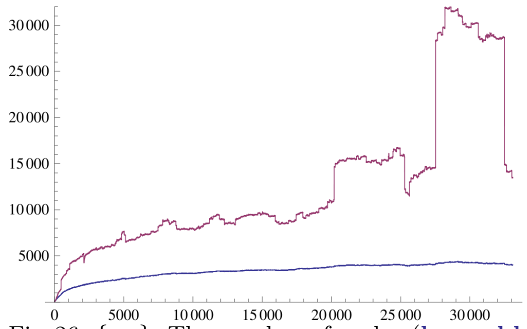

Fig. 26: {arx}: The number of nodes (**lower blue plot**) and especially the number of edges (**upper red plot**) for *Nuclear Theory* are rather volatile. Clique-wise growth proves local-unfriendly.

Various dynamics, *batch size* 1:

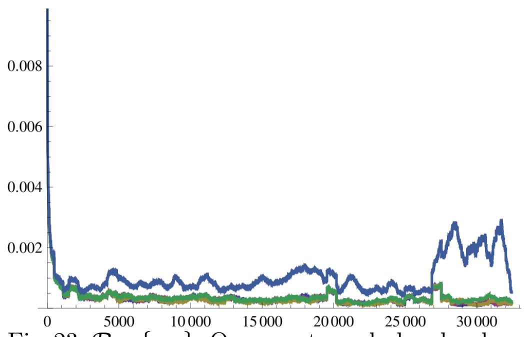

Fig. 23: $\mathcal{R}_g$, {arx}: On an extremely low level, only **dGlobal@BT** reacts to the harsh graph changes (see Fig. 26); **dLocal@N**$_1$, **dLocal@BN**$_4$, **dGlobal@BN**$_{16}$ and **dGlobal@N**$_1$ hardly spot out.

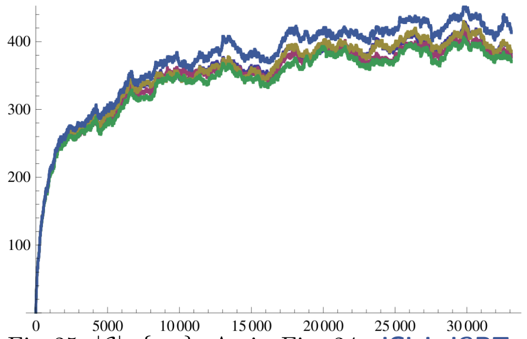

Fig. 25: $|\mathcal{C}|$, {arx}: As in Fig. 24, **dGlobal@BT** rebalances for higher quality; the locals (**dLocal@N**$_1$, **dLocal@BN**$_4$) do slightly better than the globals (**dGlobal@BN**$_{16}$, **dGlobal@N**$_1$).

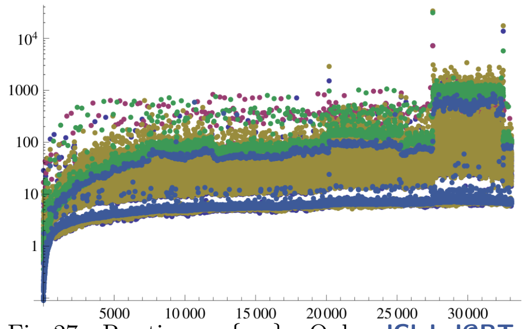

Fig. 27: Runtimes, {arx}: Only **dGlobal@BT** is largely unaffected by graph growth. Prep strategy N (**dLocal@N**$_1$, **dGlobal@N**$_1$, is slower than BN (**dLocal@BN**$_4$, **dGlobal@BN**$_{16}$).

E.3 Trials with arXiv Data: Category *Computer Science*

Category CS with all subcategories consists of 14K e-prints, 25K authors. We use *batch size* 10 for dynamics, compared to statics using *batch size* 100 in Fig. 33. dGlobal@BT excels.

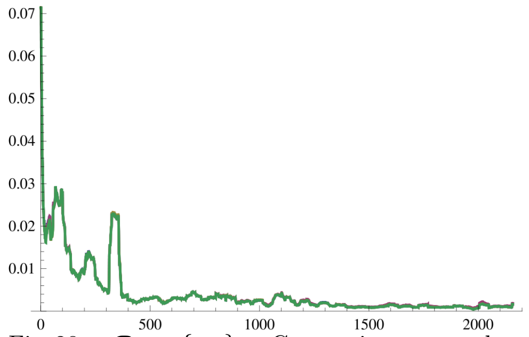

Fig. 28: $\mathcal{R}_g$, {arx}: Concerning smoothness, all tested dynamic algorithms (**dGlobal@BT**, **dLocal@BN**$_4$, **dLocal@N**$_1$, **dGlobal@BN**$_{16}$) behave almost identically, even **dGlobal@BT**.

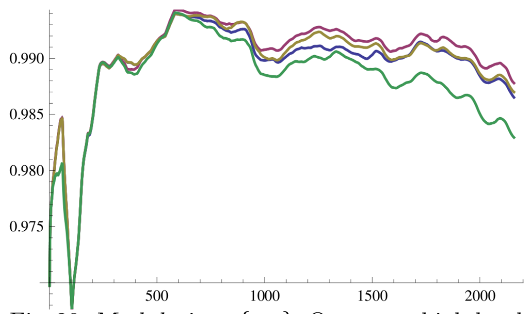

Fig. 29: *Modularity* , {arx}: On a very high level of quality, only **dGlobal@BT** starts to slightly stand out after a while, and **dGlobal@BN**$_{16}$ falls behind; **dLocal@BN**$_4$ and **dLocal@N**$_1$ hardly differ.

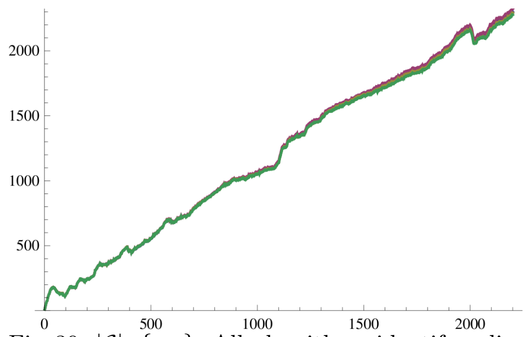

Fig. 30: $|\mathcal{C}|$, {arx}: All algorithms identify a linear growth in $|\mathcal{C}|$ over time; with **dGlobal@BT** slightly beyond the others (**dLocal@BN**$_4$, **dLocal@N**$_1$, **dGlobal@BN**$_{16}$).

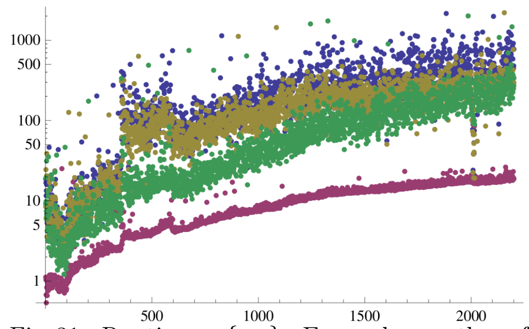

Fig. 31: Runtimes, {arx}: Even clearer than for *Nuclear Theory*, **dGlobal@BT** scales well with the graph while the other algorithms slow down more strongly (**dLocal@BN**$_4$, **dLocal@N**$_1$, **dGlobal@BN**$_{16}$).

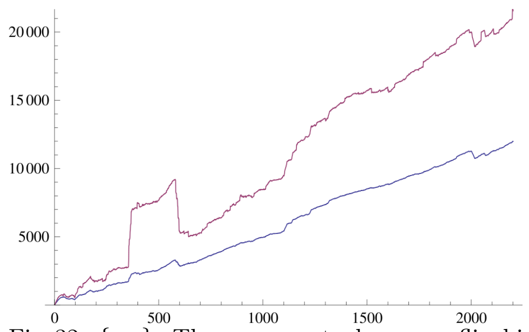

Fig. 32: {arx}: There seems to be an unflinching growth in popularity of *arXiv's* categories of computer science. Both the number of nodes (**lower blue plot**) and the number of edges (**upper red plot**) grow sharply and steadily.

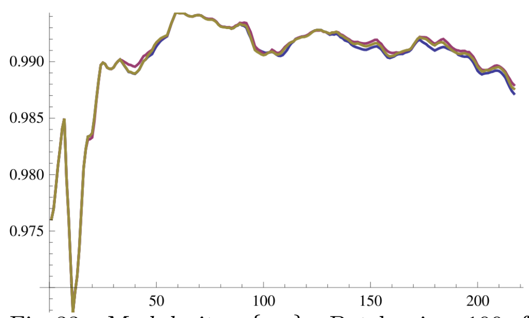

Fig. 33: *Modularity*, {arx}: Batch size 100 for statics vs. **dGlobal@BT**. By margins of 0.01% **dGlobal@BT** lies between **sLocal** and **sGlobal**. **dGlobal@BT** has virtually the same $\mathcal{R}_g$ and $|\mathcal{C}|$ as the statics but is faster by factors beyond 10^4.