# Graph Clustering: Block-models and model free results

**Authored by:**

Marina Meila
Yali Wan

**Abstract**

Clustering graphs under the Stochastic Block Model (SBM) and extensions are well studied. Guarantees of correctness exist under the assumption that the data is sampled from a model. In this paper, we propose a framework, in which we obtain "correctness" guarantees without assuming the data comes from a model. The guarantees we obtain depend instead on the statistics of the data that can be checked. We also show that this framework ties in with the existing model-based framework, and that we can exploit results in model-based recovery, as well as strengthen the results existing in that area of research.

## 1 Paper Body

: a framework for clustering with guarantees without model assumptions

In the last few years, model-based clustering in networks has witnessed spectacular progress. At the central of intact are the so-called block-models, the Stochastic Block Model (SBM), DegreeCorrected SBM (DC-SBM) and Preference Frame Model (PFM). The understanding of these models has been advanced, especially in understanding the conditions when recovery of the true clustering is possible with small or no error. The algorithms for recovery with guarantees have also been improved. However, the impact of the above results is limited by the assumption that the observed data comes from the model. This paper proposes a framework to provide theoretical guarantees for the results of model based clustering algorithms, without making any assumption about the data generating process. To describe the idea, we need some notation. Assume that a graph G on n nodes is observed. A modelbased algorithm clusters G, and outputs clustering C and parameters M(G, C).

The framework is as follows: if M(G, C) ?ts the data G well, then we shall prove that any other clustering C ? of G that also ?ts G well will be a small perturbation of C. If this holds, then C with model parameters M(G, C) can be said to capture the data structure in a meaningful way. We exemplify our approach

by obtaining model-free guarantees for the SBM and PFM models. Moreover, we show that model-free and model-based results are intimately connected.

2

Background: graphs, clusterings and block models

Graphs, degrees, Laplacian, and clustering Let G be a graph on n nodes, described by its ad? De?ne d?i = ?n A?ij the degree of node i, and D ? = diag{d?i } the diagonal jacency matrix A. j=1 ?=D ? ?1/2 A?D ? ?1/2 . In matrix of the node degrees. The (normalized) Laplacian of G is de?ned as1 L 1

? [10]. Rigorously speaking, the normalized graph Laplacian is I ? L

30th Conference on Neural Information Processing Systems (NIPS 2016), Barcelona, Spain.

extension, we de?ne the degree matrix D and the Laplacian L associated to any matrix A ? Rn?n , with Aij = Aji ? 0, in a similar way.

Let C be a partitioning (clustering) of the nodes of G into K clusters. We use the shorthand notation i ? k for ?node i belongs to cluster k?. We will represent C by its n ? K indicator matrix Z, de?ned by Zik = 1 if i ? k, 0 otherwise, for i = 1, . . . n, k = 1, . . . K. (1) T T ? Note that Z Z = diag{nk } with nk counting the number of nodes in cluster k, and Z AZ = [nkl ]K k,l=1 with nkl counting the edges in G between clusters k and l. Moreover, for two indicator matrices Z, Z ? for clusterings C, C ? , (Z T Z ? )kk? counts the number of points in the intersection of ? ? ? )kk? computes ? cluster k of C with cluster k ? of C ? , and (Z T DZ i?k?k? di the volume of the same intersection.

?Block models? for random graphs (SBM, DC-SBM, PFM) This family of models contains Stochastic Block Models (SBM) [1, 18], Degree-Corrected SBM (DC-SBM) [17] and Preference Frame Models (PFM) [20]. Under each of these model families, a graph G with adjacency matrix A? over n nodes is generated by sampling its edges independently following the law A?ij ? Bernoulli(Aij ), for all i ¿ j. The symmetric matrix A = [Aij ] describing the graph is the edge probability matrix. The three model families differ in the constraints they put on an acceptable A. Let C ? be a clustering. The entries of A are de?ned w.r.t C ? as follows (and we say that A is compatible with C ? ). SBM : Aij = Bkl whenever i ? k, j ? l, with B = [Bkl ] ? RK?K symmetric and nonnegative. DC-SBM : Aij = wi wj Bkl whenever i ? k, j ? l, with B as above and w1 , . . . wn non-negative weights associated with the graph nodes. PFM : A satis?es D = diag(A1), D?1 AZ = ZR where 1 denotes the vector of all ones, Z is the indicator matrix of C ? , and R is a stochastic matrix (R1 = 1, Rkl ? 0), the details are in [20] While perhaps not immediately obvious, the SBM is a subclass of the DC-SBM, and the latter a subclass of the PFM. Another common feature of block-models, that will be signi?cant throughout this work is that for all three, Spectral Clustering algorithms [15] have been proved to work well estimating C ? .

3

Main theorem: blueprint and results for PFM, SBM

Let M be a model class, such as SBM, DC-SBM, PFM, and denote M(G, C) ? M to be a model that is compatible with C and is ?tted in some way to graph G (we do not assume in general that this ?t is optimal). Theorem 1 (Generic

Theorem) We say that clustering C ?ts G well w.r.t M iff M(G, C) is ?close to? G. If C ?ts G well w.r.t M, then (subject to other technical conditions) any other clustering C ? which also ?ts G well is close to C, i.e. dist(C, C ? ) is small. In what follows, we will instantiate this Generic Theorem, and the concepts therein; in particular the following will be formally de?ned. (1) Model construction, i.e an algorithm to ?t a model in M to (G, C). This is necessary since we want our results to be computable in practice. (2) A goodness of ?t measure between M(C, G) and the data G. (3) A distance between clusterings. We adopt the widely used Misclassi?cation Error (or Hamming) distance de?ned below. The Misclassi?cation Error (ME) distance between two clusterings C, C ? over the same set of n points is ? 1 dist(C, C ? ) = 1 ? max 1, (2) n ??SK i?k??(k)

where ? ranges over all permutations of K elements SK , and ?(k) indexes a cluster in C ? . If the points are weighted by their degrees, a natural measure on the node set, the Weighted ME (wME) 2

distance is

? 1 distd?(C, C ? ) = 1 ? ?n max d?i . ? ??SK i=1 di i?k??(k)

(3)

? In the above, i?k?k? d?i represents the total weight of the set of points assigned to cluster k by C and to cluster k ? ( or ?(k)) by C ? . Note that in the indicator matrix representation of clusterings, this ? ? ? RK?K . While dist is more popular, we believe dist ? is the (k, k ? ) element of the matrix Z T DZ d is more natural, especially when node degrees are dissimilar, as d? can be seen as a natural measure on the set of nodes, and distd? is equivalent to the earth-mover?s distance. 3.1

Main result for PFM

Constructing a model Given a graph G and a clustering C of its nodes, we wish to construct a ? ? L—— is small. PFM compatible with C, so that its Laplacian L satis?es that ——L

? be Let the spectral decomposition of L ?? ? ? ? Y? T ? 0 T ? low Y?low ? Y? T + Y?low ? ? ? ? = Y? ? L = [Y Ylow ] (4) T ? low Y?low 0 ? ?1, ? ? ? , ? ? K ), ? ? K+1 , ? ? ? , ? ? n ). To ? = diag(? ? low = diag(? where Y? ? Rn?K , Y?low ? Rn?(n?K) , ? ? ? ? ensure that the matrices Y , Ylow are uniquely de?ned we assume throughout the paper that L?s K-th eigengap, i.e, —?K — ? —?K+1 —, is non-zero. ? 2 — ? . . . ? —? ? K — ¿ —? ? K+1 — ? . . . —? ? n —. ? 1 = 1 ? —? ? satisfy ? Assumption 1 The eigenvalues of L

Denote the subspace spanned by the columns of M , for any M matrix, by R(M ), and —— —— the Euclidean or spectral norm. PFM Estimation Algorithm ? D, ? L, ? Y? , ?, ? clustering C with indicator matrix Z. Input Graph G with A, Output (A, L) = P F M (G, C) 1. Construct an orthogonal matrix derived from Z. ? the column normalization of Z. ? 1/2 ZC ?1/2 , with C = Z TDZ YZ = D ? Note Ckk = i?k d?i is the volume of cluster k.

(5)

2. Project YZ on Y? and perform Singular Value Decomposition. F = YZT Y? = U ?V T

(6)

3. Change basis in R(YZ ) to align with Y? . Y = YZ U V T .

Complete Y to an orthonormal basis [Y B] of Rn .

(7)

4. Construct Laplacian L and edge probability matrix A. T ? T + (BB T )L(BB ? L = Y ?Y ),

? 1/2 LD ? 1/2 . A = D

(8)

? D, ? L, ? Y? , ? ? and Z be de?ned as above, and (A, L) = P F M (G, C). Then, Proposition 2 Let G, A, ? and L, or A de?ne a PFM with degrees d?1:n . 1. D ? 1:K . 2. The columns of Y are eigenvectors of L with eigenvalues ? ? 1 = 1. ? 1/2 1 is an eigenvector of both L and L ? with eigenvalue ? 3. D The proof is relegated to the Supplement, as are all the omitted proofs. P F M (G, C) is an estimator for the PFM parameters given the clustering. It is evidently not the Maximum Likelihood estimator, but we can show that it is consistent in the following sense. 3

Proposition 3 (Informal) Assume that G is sampled from a PFM with parameters D? , L? and compatible with C ? , and let L = P F M (G, C ? ). Then, under standard recovery conditions for PFM (e.g [20]) ——L? ? L—— = o(1) w.r.t. n. ? ? L—— ? ?. Assumption 2 (Goodness of ?t for PFM) ——L P F M (G, C) instantiates M(G, C), and Assumption 2 instantiates the goodness of ?t measure. It remains to prove an instance of Generic Theorem 1 for these choices. ? 1:n as de?ned, and L ? L, ? ? ? satTheorem 4 (Main Result (PFM)) Let G be a graph with d?1:n , D, isfy Assumption 1. Let C, C ? be two clusterings with K clusters, and L, L? be their corresponding 2 and Laplacians, de?ned as in (8), and satisfy Assumption 2 respectively. Set ? = (—?? (K?1)? ? —?—? —)2 K

K+1

?0 = mink Ckk / maxk Ckk with C de?ned as in (5), where k indexes the clusters of C. Then, whenever ? ? ?0 , maxk Ckk ?, distd?(C, C ? ) ? ? k Ckk

(9)

with distd? being the weighted ME distance (3).

In the remainder of this section we outline the proof steps, while the partial results of Proposition 5, 6, 7 are proved in the Supplement. First, we apply the perturbation bound called the Sinus Theorem of Davis and Kahan, in the form presented in Chapter V of [19]. ? 1:n , Y be de?ned as usual. If Assumptions 1 and 2 hold, then Proposition 5 Let Y? , ? —— diag(sin ?1:K (Y? , Y ))—— ?

? = ?? ? ? K+1 — — ?K — ? — ?

(10)

where ?1:K are the canonical (or principal) angles between R(Y? ) and R(Y ) (see e.g [8]). The next step concerns the closeness of Y, Y? in Frobenius norm. Since Proposition 5 bounds the sinuses of the canonical angles, we exploit the fact that the cosines of the same angles are the singular values of F = Y T Y? of (6). ? = Y? Y? T and F, ?? as above. Assumptions 1 and 2 imply that Proposition 6 Let M = Y Y T , M ? T ? K ? (K ? 1)??2 . 1. ——F ——2F = trace M M ? ——2 ? 2(K ? 1)??2 . 2. ——M ? M F Now we show that

all clusterings which satisfy Proposition 6 must be close to each other in the weighted ME distance. For this, we ?rst need an intermediate result. Assume we have two clusterings C, C ? , with K clusters, for which we construct YZ , Y, L, M , respectively YZ? , Y ? , L? , M ? as above. Then, the subspaces spanned by Y and Y ? will be close. ? satisfy Assumption 1 and let C, C ? represent two clusterings for which L, L? Proposition 7 Let L satisfy Assumption 2. Then, ——YZT YZ? ——2F ? K ? 4(K ? 1)??2 = K ? ? The main result now follows from Proposition 7 and Theorem 9 of [13], as shown in the Supplement. This proof approach is different from the existing perturbation bounds for clustering, which all use counting arguments. The result of [13] is a local equivalence, which bounds the error we need in terms of ? de?ned above (?local? meaning the result only holds for small ?). 4

3.2

Main Theorem for SBM

In this section, we offer an instantiation of Generic Theorem 1 for the case of the SBM. As before, we start with a model estimator, which in this case is the Maximum Likelihood estimator. SBM Estimation Algorithm ? clustering C with indicator matrix Z. Input Graph with A, Output A = SBM (G, C)

1. Construct an orthogonal matrix derived from Z: YZ = ZC ?1/2 with C = Z T Z. ?1 ? 2. Estimate the edge probabilities: B = C ?1 Z T AZC . 3. Construct A from B by A = ZBZ T . ? = C 1/2 BC 1/2 and denote the eigenvalues of B, ? ordered by decreasing magProposition 8 Let B ? be B ? = U ?U T , with U an orthogonal matrix nitude, by ?1:K . Let the spectral decomposition of B and ? = diag(?1:K ). Then 1. A is a SBM. 2. ?1:K are the K principal eigenvalues of A. The remaining eigenvalues of A are zero. 3. A = Y ?Y T where Y = YZ U . Assumption 3 (Eigengap) B is non-singular (or, equivalently, —?K — ¿ 0. Assumption 4 (Goodness of ?t for SBM) ——A? ? A—— ? ?. With the model (SBM), estimator, and goodness of ?t de?ned, we are ready for the main result. ? A the K-th ? and ? Theorem 9 (Main Result (SBM)) Let G be a graph with incidence matrix A, K ? ? singular value of A. Let C, C be two clusterings with K clusters, satisfying Assumptions 3 and 4. 2 Set ? = —4K? ? A —2 and ?0 = mink nk / maxk nk , where k indexes the clusters of C. Then, whenever ? K

? ? ?0 , dist(C, C ? ) ? ? maxk nk /n, where dist represents the ME distance (2).

? ? ? A is not bounded above, and neither is ?. Since the SBM is less Note that the eigengap of A, K ?exible than the PFM, we expect that for the same data G, Theorem 9 will be more restrictive than Theorem 4.

4 4.1

The results in perspective Cluster validation

Theorems like 4, 9 can provide model free guarantees for clustering. We exemplify this procedure in the experimental Section 6, using standard spectral clustering as described in e.g [18, 17, 15]. What is essential is that all the quantities such as ? and ? are computable from the data. Moreover, if Y is available, then the bound in Theorem 4 can be improved. ? , M ¿F +(K ? 1)(?? )2 + Proposition 10 Theorem 4 holds when ? is replaced by ?Y = K? ¡ M ? ?

? ? ? ? ? 2 2(K ? 1)? ——M ? M ——F , with ? = ?/(—?K — ? —?K+1 —) and M, M de?ned in Proposition 6. 4.2

Using existing model-based recovery theorems to prove model-free guarantees

We exemplify this by using (the proof of) Theorem 3 of [20] to prove the following. Theorem 11 (Alternative result based on [20] for PFM) Under the same conditions as in TheoK?2 . rem 4, distd?(C, C ? ) ? ?WM , with ?WM = 128 (—?? —?— ? ? —)2 K

5

K+1

It follows, too, that with the techniques in this paper, the error bound in [20] can be improved by a factor of 128. Similarly, if we use the results of [18] we obtain alternative model-free guarantee for the SBM. ? 2 ?L2 ——F ? ?, where L, ? L are the LaplaAssumption 5 (Alternative goodness of ?t for SBM) ——L ? cians of A and A = SBM (G, C) respectively. Theorem 12 (Alternative result based on [18] for SBM) Under the same conditions as in Theo2 k nk . rem 9, except for replacing Assumption 4 with 5, dist(C, C ? ) ? ?RCY with ?RCY = —??? —4 16 max n K

A problem with this result is that Assumption 5 is much stronger than 4 (being in Frobenius norm). The more recent results of [17] (with unspeci?ed constants) in conjunction with our original As?2 ) sumptions 3, 4, and the assumption that all clusters have equal sizes, give a bound of O(K?2 /? K for the SBM; hence our model-free Theorem 9 matches this more restrictive model-based theorem. 4.3

Sanity checks and Extensions

It can be easily veri?ed that if indeed G is sampled from a SBM, or PFM, then for large enough n, and large enough model eigengap, Assumptions 1 and 2 (or 3 and 4) will hold. Some immediate extensions and variations of Theorems 4, 9 are possible. For example, one could replace the spectral norm by the Frobenius norm in Assumptions 2 and 4, which would simplify some of the proofs. However, using the Frobenius norm would be a much stronger assumption [18] Theorem 4 holds not just for simple graphs, but in the more general case when A? is a weighted graph (i.e. a similarity matrix). The theorems can be extended to cover the case when C ? is a clustering ? ? ——L ? L——(1 ? that is ?-worse than C, i.e when ——L? ? L—— ? ?). 4.4

Clusterability and resilience

?? Our Theorems also imply the stability of a clustering to perturbations of the graph G. Indeed, let L ? ? ? ? ? ? ? be the Laplacian of G , a perturbation of G. If ——L ? L—— ? ?, then ——L ? L—— ? 2?, and (1) G is well ?tted by a PFM whenever G is, and (2) C is ? stable w.r.t G ? , hence C is what some authors [9] call resilient. A graph G is clusterable when G can be ?tted well by some clustering C ? . Much work [4, 7] has been devoted to showing that clusterability implies that ?nding a C close to C ? is computationally ef?cient. Such results can be obtained in our framework, by exploiting existing recovery theorems such as [18, 17, 20], which give recovery guarantees for Spectral Clustering, under the assumption of sampling from the model. For this, we can

simply replace the model assumption with the assumption that there is a C ? for which L (or A) satis?es Assumptions 1 and 2 (or 3 and 4).

5

Related work

To our knowledge, there is no work of the type of Theorem 1 in the literature on SBM, DC-SBM, PFM. The closest work is by [6] which guarantees approximate recovery assuming G is close to a DC-SBM. Spectral clustering is also used for loss-based clustering in (weighted) graphs and some stability results exist in this context. Even though they measure clustering quality by different criteria, so that the ? values are not comparable, we review them here. The recent paper of [16], Theorem 1.2 states ? K+1 )/(cK 3 ) then the clustering error2 that if the K-way Cheeger constant of G is ?(k) ? (1 ? ? opt distd?(C, C ) ? C/c = ?P SZ . In the current proof, the constant C = 2 ? 105 ; moreover, ?(K) cannot be computed tractably. In [14], the bound ?M SX depends on ?M SX , the Normalized Cut scaled by the eigengap. Since both bounds refer to the result of spectral clustering, we can compare the relationship between ?M SX and ?M SX ; for [14], this is ?M SX = 2?M SX [1 ? ?M SX /(K ? 1)], 2 The results is stronger, bounding the perturbation of each cluster individually by ?P SZ , but it also includes a factor larger than 1, bounding the error of K-means algorithm.

6

which is about K ? 1 times larger than ? when ? = ?M SX . In [5], dist(C, C ? ) is de?ned in terms of ——YZT ? YZ? ——2F , and the loss is (closely related) to ——A? ? SBM (G, C)——2F . The bound does not take into account the eigengap, that is, the stability of the subspace Y? itself. Bootstrap for validating a clustering C was studied in [11] (see also references therein for earlier work). In [3] the idea is to introduce a statistics, and large deviation bounds for it, conditioned on sampling from a SBM (with covariates) and on a given C.

6

Experimental evaluation

Experiment Setup Given G, we obtain a clustering C0 by spectral clustering [15]. Then we calculate clustering C by perturbing C0 with gradually increasing noise. For each C, we construct PFM (C, G)and SBM(C, G) model, and further compute ?, ? and ?0 . If ? ? ?0 , C is guaranteed to be stable by the theorems. In the remainder of this section, we describe the data generating process for the simulated datasets and the results we obtained. PFM Datasets We generate from PFM model with K = 5, n = 10000, ?1:K = (1, 0.875, 0.75, 0.625, 0.5). eigengap = 0.48, n1:K = (2000, 2000, 2000, 2000, 2000). The stochastic matrix R and its stationary distribution ? are shown below. We sample an adjacency matrix A? from A (shown below).

A ? 25 .12 .17 .18 .28 .79 .02 .06 .03 .10 .71 .23 .00 .02 ? .03 ? .16 .69 .00 .06 R = ? .09 ? .04 .00 .00 .80 .16 .10 .01 .03 .11 .76 ?=

?

?

A?

? ? ? ? ?

Perturbed PFM Datasets A is obtained from the previous model by perturbing its principal subspace (details in Supplement). Then we sample A? from A. Lancichinetti-Fortunato-Radicchi (LFR) simulated matrix [12] The LFR benchmark graphs are widely used for community detection algorithms, due to heterogeneity in the distribution of node degree and community size. A LFR matrix is simulated with n = 10000, K = 4, nk = (2467, 2416, 2427, 2690) and ? = 0.2, where ? is the mixing parameter indicating the fraction of edges shared between a node and the other nodes from outside its community. ? of hyperlinks between weblogs on US politics, Political Blogs Dataset A directed network A compiled from online directories by Adamic and Glance [2], where each blog is assigned a political leaning, liberal or conservative, based on its blog content. The network A contains 1490 blogs. ? T A) ? 3 , which is a smoothed After erasing the disconnected nodes, n = 983. We study A? = (A T ? ? undirected graph. For A A we ?nd no guarantees. The ?rst two data sets are expected to ?t the PFM well, but not the SBM, while the LFR data is expected to be a good ?t for a SBM. Since all bounds can be computed on weighted graphs as well, we have run the experiments also on the edge probability matrices A used to generate the PFM and perturbed PFM graphs. The results of these experiments are summarized in Figure 1. For all of the experiments, the clustering C is ensured to be stable by Theorem 4 as the unweighted error grows to a breaking point, then the assumptions of the theorem fail. In particular, the C0 is always stable in the PFM framework. 7

Comparing ? from Theorem 9 to that from Theorem 4, we ?nd that Theorem 9 (guarantees for SBM) is much harder to satisfy. All ? values from Theorem 9 are above 1, and not shown.3 In particular, for the SBM model class, the C cannot be proved stable even for the LFR data.

Note that part of the reason why with the PFM model very little difference from the clustering C0 can be tolerated for a clustering to be stable is that the large eigengap makes P F M (G, C) differ from P F M (G, C0 ) even for very small perturbations. By comparing the bounds for A? with the bounds for the ?weighted graphs? A, we can evaluate that the sampling noise on ? is approximately equal to that of the clustering perturbation. Of course, the sampling noise varies with n, decreasing for larger graphs. Moreover, from Political Blogs data, we see that ?smoothing? a graph, by e.g. taking powers of its adjacency matrix, has a stability inducing effect.

Figure 1:

? denotes a simple graph, while A denotes a Quantities ?, ?, ?0 from Theorem 4 plotted vs dist(C, C0 ) for various datasets: A

weighted graph (i.e. a non-negative matrix). For the Political Blogs: Truth means C0 is true clustering of [2], spectral means C0 is obtained

from spectral clustering. For SBM, ? is always greater than ?0 .

7

Discussion

This paper makes several contributions. At a high level, it poses the problem of model free validation in the area of community detection in networks. The stability paradigm is not entirely new, but using it explicitly with model-based

clustering (instead of cost-based) is. So is ?turning around? the model-based recovery theorems to be used in a model-free framework. All quantities in our theorems are computable from the data and the clustering C, i.e do not contain undetermined constants, and do not depend on parameters that are not available. As with distribution-free results in general, making fewer assumptions allows for less con?dence in the conclusions, and the results are not always informative. Sometimes this should be so, e.g when the data does not ?t the model well. But it is also possible that the ?t is good, but not good enough to satisfy the conditions of the theorems as they are currently formulated. This happens with the SBM bounds, and we believe tighter bounds are possible for this model. It would be particularly interesting to study the non-spectral, sharp thresholds of [1] from the point of view of model-free recovery. A complementary problem is to obtain negative guarantees (i.e that C is not unique up to perturbations). At the technical level, we obtain several different and model-speci?c stability results, that bound the perturbation of a clustering by the perturbation of a model. They can be used both in model-free and in existing or future model-based recovery guarantees, as we have shown in Section 3 and in the experiments. The proof techniques that lead to these results are actually simpler, more direct, and more elementary than the ones found in previous papers. 3

We also computed ?RCY but the bounds were not informative.

8

# 2 References

[1] Emmanuel Abbe and Colin Sandon. Community detection in general stochastic block models: fundamental limits and ef?cient recovery algorithms. arXiv preprint arXiv:1503.00609, 2015. [2] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 us election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery, pages 36?43. ACM, 2005. [3] Edoardo M. Airoldi, David S. Choi, and Patrick J. Wolfe. Con?dence sets for network structure. Technical Report arXiv:1105.6245, 2011. [4] Pranjal Awasthi. Clustering under stability assumptions. In Encyclopedia of Algorithms, pages 331?335. 2016. [5] Francis Bach and Michael I. Jordan. Learning spectral clustering with applications to speech separation. Journal of Machine Learning Research, 7:1963?2001, 2006. [6] Maria-Florina Balcan, Christian Borgs, Mark Braverman, Jennifer Chayes, and Shang-Hua Teng. Finding endogenously formed communities. In Proceedings of the Twenty-Fourth Annual ACM-SIAM Symposium on Discrete Algorithms, pages 767?783. SIAM, 2013. [7] Shai Ben-David. Computational feasibility of clustering under clusterability assumptions. CoRR, abs/1501.00437, 2015. [8] Rajendra Bhatia. Matrix analysis, volume 169. Springer Science & Business Media, 2013. [9] Yonatan Bilu and Nathan Linial. Are stable instances easy? CoRR, abs/0906.3162, 2009. [10] Fan RK Chung. Spectral graph theory, volume 92. American Mathematical Soc., 1997. [11] Brian Karrer, Elizaveta Levina, and M. E. J. Newman. Robustness of community structure in networks. Phys. Rev. E, 77:046119, Apr 2008. [12] An-

drea Lancichinetti, Santo Fortunato, and Filippo Radicchi. Benchmark graphs for testing community detection algorithms. Physical review E, 78(4):046110, 2008. [13] Marina Meil?a. Local equivalence of distances between clusterings ? a geometric perspective. Machine Learning, 86(3):369?389, 2012. [14] Marina Meil?a, Susan Shortreed, and Liang Xu. Regularized spectral learning. In Robert Cowell and Zoubin Ghahramani, editors, Proceedings of the Arti?cial Intelligence and Statistics Workshop(AISTATS 05), 2005. [15] A. Y. Ng, M. I. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In T. G. Dietterich, S. Becker, and Z. Ghahramani, editors, Advances in Neural Information Processing Systems 14, Cambridge, MA, 2002. MIT Press. [16] Richard Peng, He Sun, and Luca Zanetti. Partitioning well-clustered graphs with k-means and heat kernel. In Proceedings of the Annual Conference on Learning Theory (COLT), pages 1423?1455, 2015. [17] Tai Qin and Karl Rohe. Regularized spectral clustering under the degree-corrected stochastic blockmodel. In Advances in Neural Information Processing Systems, pages 3120?3128, 2013. [18] Karl Rohe, Sourav Chatterjee, and Bin Yu. Spectral clustering and the high-dimensional stochastic blockmodel. The Annals of Statistics, pages 1878?1915, 2011. [19] Gilbert W Stewart and Ji-guang Sun. Matrix perturbation theory, volume 175. Academic press New York, 1990. [20] Yali Wan and Marina Meila. A class of network models recoverable by spectral clustering. In Daniel Lee and Masashi Sugiyama, editors, Advances in Neural Information Processing Systems (NIPS), page (to appear), 2015.

9