Reflexive Regular Equivalence for Bipartite Data

Aaron Gerow[1*], Mingyang Zhou[2], Stan Matwin[1], and Feng Shi[3]

[1] Faculty of Computer Science, Dalhousie University, Halifax, NS, Canada
[2] Department of Computer Science, University of Chicago, Chicago, USA
[3] University of North Carolina, Chapel Hill, NS, USA
*Corresponding author: gerow@dal.ca

Abstract. Bipartite data is common in data engineering and brings unique challenges, particularly when it comes to clustering tasks that impose strong structural assumptions. This work presents an unsupervised method for assessing similarity in bipartite data. The method is based on *regular equivalence* in graphs and uses spectral properties of a bipartite adjacency matrix to estimate similarity in both dimensions. The method is reflexive in that similarity in one dimension informs similarity in the other. The method also uses local graph transitivities, a contribution governed by its only free parameter. Reflexive regular equivalence can be used to validate assumptions of co-similarity, which are required but often untested in co-clustering analyses. The method is robust to noise and asymmetric data, making it particularly suited for cluster analysis and recommendation in data of unknown structure.[4]

In bipartite data, co-similarity is the notion that similarity in one dimension is matched by similarity in some other dimension. Such data occurs in a many areas: text mining, gene expression networks, consumer co-purchasing data and social affiliation. In bipartite analyses, co-clustering is an increasingly prominent technique in a range of applications [11], but has strong co-similarity assumptions. One example of co-similar structure is in text analysis where similar words appear in similar documents, where there is assumed to be a permutation of the word-document co-occurrence matrix that exposes co-similarity among words and documents [4]. The work here describes a way to *assess* co-similarity using regular equivalence [8] with a reflexive conception of similarity that accommodates nodes' (data-points) local structures. This assessment is a kind of pre-condition for co-clustering: if there is little co-similarity, co-clustering will yield a poor clustering solution. Assessing co-similarity will produce similarity in one dimension to expose potential clustering without requiring it across dimensions. This is particularly useful for asymmetric data when a non-clustered dimension informs, but does not reciprocate clustering in the other. Whereas as co-clustering finds clusters across dimensions, our method provides a decoupled solution in each mode. Reflexive regular equivalence is able to quantify how much one dimension informs similarity in the other. Additionally, our results show that by incorporating local structures, it can better overcome noise and accommodate asymmetry.

[4] An extended preprint of this paper is available at arxiv.org/abs/1702.04956.

Background Measuring co-similarity amounts to calculating similarity in two dimensions. In a graph setting, this can be done with bipartite generalizations of regular equivalence [1,5], which measure similarity based on the similarity of neighboring nodes. Bipartite regular equivalence makes sense of *within-dimension regularity* and *between-dimension structure*. As we will see, allowing the inter-dimensional structure to be "reflexive", as defined below, greatly helps assess co-similarity. Specifically, we conceive of bipartite data as a two-mode network. In simple networks, vertex similarity can be measured using pairwise metrics like the Jaccard index, the cosine of a pair's connectivity patterns, their correlation or simply the overlap of neighbors. Recent efforts have sought to account for structure beyond nodes' immediate neighbors. *Regular equivalence* proposes that nodes are similar to the extent their neighbors are similar [8], a method common in social network analysis [7].

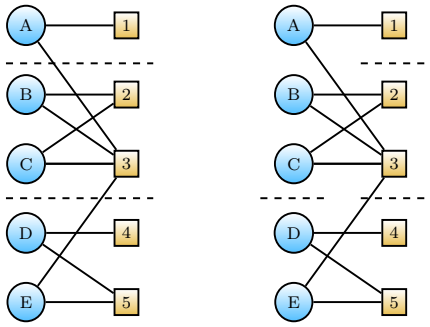

Fig. 1. With the same network (left; a) a biclustering solution yields partitions that group vertices in both modes. On the right (b) is a hypothetical grouping based on similarities produced by reflexive regular equivalence; there may not be a mapping between clusters in each mode and groupings are exposed by similarity, not by partitioning.

In a social setting, co-similarity was formalized by Ronald Breiger in 1974 where similar people were found to participate in similar groups [2]. With the adjacency structure, A, of $people \times groups$, similarity can be measured as $S_{people} = AA^{\mathsf{T}}$ and $S_{groups} = A^{\mathsf{T}}A$. Such multi-modal conceptions of social structures are now common in models of social search and recommendation [6,10]. Our method extends regular equivalence and Breiger's notion of co-similarity. Crucially, it considers structural equivalence by adding contributions from local transitivity – two nodes are similar if their neighbors are similar within and across dimensions. Our goal is not to find partitions that optimally group nodes across dimensions, but to measure similarity in each mode of the data using information from the other. Only by extension, may any clustered structure be revealed in the data. Figure 1 depicts a bipartite network and hypothetical co-clustering solution (a) and a potential solution based on reflexive regular equivalence (b). Though different from our method, co-clustering algorithms have much in common with reflexive regular equivalence. Co-clustering algorithms yield partitions across modes that group self-similar rows with self-similar columns, and assume co-similarity just as clustering assumes similarity. Reflexive equivalence assess

co-similarity, which *may* justify a co-clustering strategy. For examples of co-clustering algorithms that implement related notions of similarity see [3,9,12].

Method Our reflexive equivalence method is an unsupervised approach to measuring vertex similarity in a bipartite network. This notion is equivalent to a pairwise similarity metric that operates on row- and column-vectors of the adjacency structure in bimodal data. Bipartite networks are equivalent to what are normally treated as rows and columns in a matrix. Our method iterates between each mode incorporating information from the other. There is no restriction to binary-valued data: edges may be weighted.

Let G be a bipartite graph with two sets of nodes, V and V' and A be the adjacency matrix in which $A_{ij} > 0$ represents how strongly $i \in V$ is connected to $j \in V'$. Let S be a square $|V|$ matrix where entry S_{ij} is the similarity between i and j in V, and S' be a square $|V'|$ matrix where S'_{ij} is the similarity between i and j in V'. Assuming the similarity of i and j in one mode is informed by similarity between neighbors of i and any neighbor of j in the other mode, then

$$S_{ij} = \sum_k \sum_l A_{ik} A_{jl} S'_{kl}, \tag{1}$$

$$S'_{ij} = \sum_k \sum_l A_{ki} A_{lj} S_{kl}, \tag{2}$$

This provides a procedure for inferring similar nodes in the adjacency matrix. Starting from randomly initialized S and S', Eqs. 1 and 2 can be applied iteratively until convergence of $\|S\|_{\mathsf{F}}$ and $\|S'\|_{\mathsf{F}}$. This formulation, however, treats all neighbors equally, regardless of their structural importance to the pair of nodes in question. Thus, we weight similarity between common neighbors more than non-common neighbors. Denoting neighbors of i and j by Γ_i and Γ_j respectively, we define this similarity as follows:

$$S_{ij} = (1-\alpha)[\sum_{k \in (\Gamma_i - \Gamma_i \cap \Gamma_j)} \sum_{l \in \Gamma_j} S'_{kl} + \sum_{k \in \Gamma_i \cap \Gamma_j} \sum_{l \in (\Gamma_j - \Gamma_i \cap \Gamma_j)} S'_{kl}] + \sum_{k \in \Gamma_i \cap \Gamma_j} \sum_{l \in \Gamma_i \cap \Gamma_j} S'_{kl},$$

$$S'_{ij} = (1-\alpha)[\sum_{k \in (\Gamma_i - \Gamma_i \cap \Gamma_j)} \sum_{l \in \Gamma_j} S_{kl} + \sum_{k \in \Gamma_i \cap \Gamma_j} \sum_{l \in (\Gamma_j - \Gamma_i \cap \Gamma_j)} S_{kl}] + \sum_{k \in \Gamma_i \cap \Gamma_j} \sum_{l \in \Gamma_i \cap \Gamma_j} S_{kl}.$$

This combines structural equivalence, regular equivalence and reflexivity into a single model. A parameter α balances the contribution of non-common and common neighbors. Rearranging the terms, the equations can be rewritten as

$$S_{ij} = (1-\alpha) \sum_k \sum_l A_{ik} A_{jl} S'_{kl} + \alpha \sum_k \sum_l A_{ik} A_{jk} A_{il} A_{jl} S'_{kl}, \tag{3}$$

$$S'_{ij} = (1-\alpha) \sum_k \sum_l A_{ki} A_{lj} S_{kl} + \alpha \sum_k \sum_l A_{ki} A_{kj} A_{li} A_{lj} S_{kl}, \tag{4}$$

or in matrix form:

$$S = (1-\alpha) A S' A^{\mathsf{T}} + \alpha (A \otimes A^{\mathsf{T}}) \cdot S' \cdot (A \otimes A^{\mathsf{T}})^{\mathsf{T}}, \tag{5}$$

$$S' = (1-\alpha) A^{\mathsf{T}} S A + \alpha (A^{\mathsf{T}} \otimes A) \cdot S \cdot (A^{\mathsf{T}} \otimes A)^{\mathsf{T}}, \tag{6}$$

where $(\cdot)$ is the conventional inner product defined in tensor algebra. Eqs. 5 and 6 compute reflexivity between the two dimensions and the effect of local structure is controlled by α. When $\alpha = 0$, the method is a bipartite form of regular equivalence. As α increases, similarity between common neighbors plays a larger role. A is normalized prior to applying the algorithm, and S and S' are normalized after every iteration by their L_1-, L_2- or L_∞-vector norm.

Results We evaluated the method on data with known structure. A set of semi-random $n \times m$ versions of A were generated with diagonal blocks of random sizes: the resulting similarity matrices should have diagonal blocks proportional to the row- and column-wise block sizes in A. The test is to run the algorithm on randomly permuted data, $\hat{A}$, after which if we apply the original ordering of A to $\hat{S}$ and $\hat{S}'$. The solution is then assessed as

$$\mu = \frac{1}{2}\|S - \hat{S}\|_{\mathsf{F}} \frac{1}{|S|} + \frac{1}{2}\|S' - \hat{S}'\|_{\mathsf{F}} \frac{1}{|S'|}. \tag{7}$$

Here, μ is simply the mean difference between results on A and $\hat{A}$, and in the perfect case will be 0. We compare reflexive similarity to three pairwise metrics operating on the rows and columns of A. The results show that reflexive variants perform marginally better (Figure 1b); α had no effect, which is expected given all node interactions are purely local.

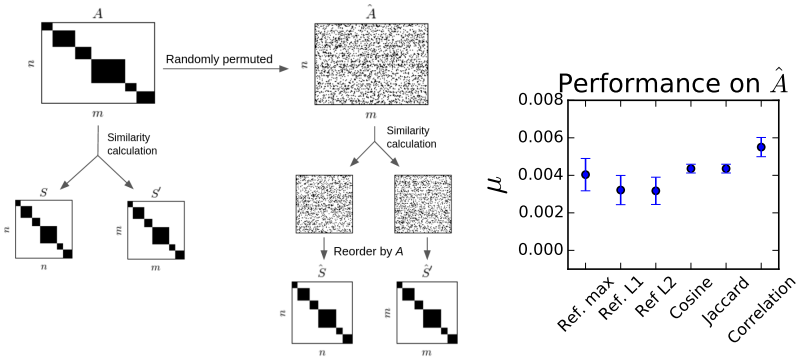

Fig. 2. (a; left) Similarity of $\hat{A}$ should produce results similar to those computed for A. (b; right) Performance (μ; lower is better) on $\hat{A}$ using variants of reflexive similarity method. Error-bars are ± 2 s.e. of the mean over ten versions of A.

Robustness to Noise Because real-world data is often noisy, we evaluated the method with noise added to A of the form $\mathcal{N}(0, \sigma)$ to produce $\tilde{A}$. Note that adding noise breaks the element-wise symmetry of A but the underlying block-structure remains symmetric. Figure 3 shows results with respect to noise, normalization variant and α. With little noise, the L_∞-norm variant outperforms other methods when $\alpha > 0.5$. As noise is increased, pairwise metrics perform worse, as do the reflexive methods with $\alpha = 0$. This confirms that our method is well-suited to finding co-similarity structure in both dimensions of noisy data where the L_∞ variant with $\alpha = 1$ performs best.

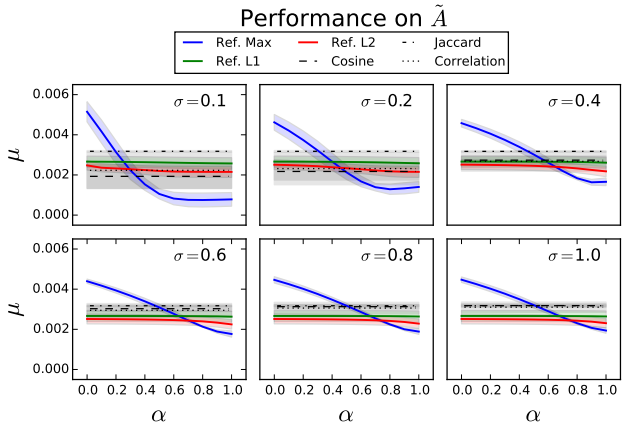

Fig. 3. Performance (μ; lower is better) with respect to noise, σ, in $\tilde{A}$ for different values of α. Error-bands are ± 2 s.e. of the mean over ten versions of A.

Unbalanced Co-similarity Because clusters exposed by similarity in S and S' are not coupled as they are in co-clustering tasks, reflexive similarity is able to find unbalanced co-similarity structure. With this kind of data, the reflexive method should produce similarity matrices with different structure. For this task, another set of semi-random versions of A were generated with random unbalanced block structure, which were then randomly permuted to get $\hat{A}$. Results on $\hat{A}$ were compared to results from A. Figure 4a shows the results on variants of reflexive regular equivalence and the pairwise metrics. Results show that all variants of reflexive similarity outperform the pairwise metrics, but that neither normalization choice nor α make significant difference.

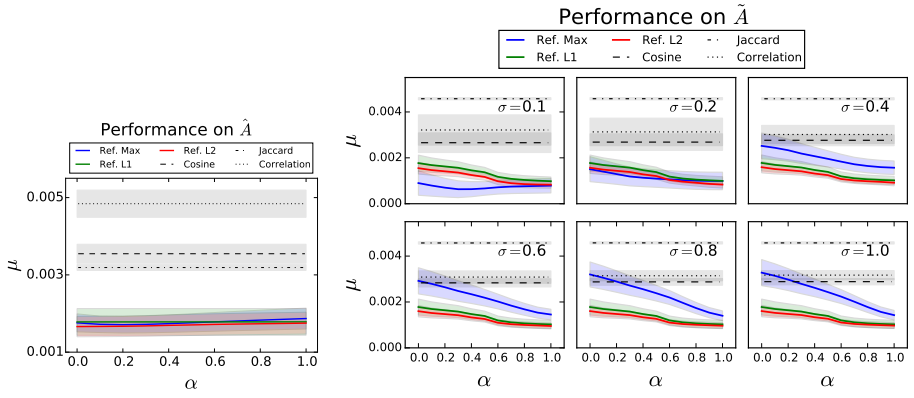

Fig. 4. (a; left) Performance on ten versions of $\hat{A}$ with unbalanced structure and, (b; left) with varying amounts of noise.

The parameter α is crucial to our method's ability to overcome noise in symmetric data. The same task was run on asymmetric data with added noise. Figure 4b shows performance with different noise levels. The results show that after adding a small amount of noise, the reflexive variants significantly outperform pairwise metrics. In particular, for low levels of noise, the L_∞-norm variant performs best, but for higher levels ($\sigma > 0.2$) the L_1- and L_2-norms are better. In all variants, higher values of α yield better performance. This suggests local structure is useful even when co-similarity is asymmetric. Overall, the evaluations show that when the structure of A is known, reflexive similarity is able to leverage inter-dimensional similarity in noisy and asymmetric data to provide better results than methods restricted to one dimension.

Discussion Co-clustering analysis often employs trial-and-error when assessing similarity. There are two problems with this: a priori estimates of the number of clusters can be hard to make, and clusters may not be coupled across dimension. Reflexive regular equivalence offers a way to attenuate inter-dimensional structure and local transitivities in similarity calculations. The results show this is particularly important in noisy data. The method also offers a way to validate co-similarity assumptions: if the same permutation for S and S' exposes block structure in each, then A is co-similar. By varying α, one can enhance or reduce the contribution of local equivalence, essentially backing off to spectral similarity. In this way, our method offers a way to measure co-similarity across dimensions, helping confirm assumptions of co-similarity.

References

1. S. P. Borgatti. 2-mode concepts in social network analysis. *Encyclopedia of complexity and system science*, 6, 2009.
2. R. L. Breiger. The duality of persons and groups. *Social forces*, 53:181–190, 1974.
3. V. Codocedo and A. Napoli. Lattice-based biclustering using partition pattern structures. In *ECAI 14'*, pages 213–218, 2014.
4. I. S. Dhillon. Co-clustering documents and words using bipartite spectral graph partitioning. In *Proc. SIGKDD 01'*, pages 269–274, 2001.
5. P. Doreian, V. Batagelj, and A. Ferligoj. Generalized blockmodeling of two-mode network data. *Social networks*, 26(1):29–53, 2004.
6. A. Gerow, B. Lou, E. Duede, and J. Evans. Proposing ties in a dense hypergraph of academics. In *Proc. of Social Informatics 2015*, pages 209–226. Springer, 2015.
7. D. Gnatyshak, D. I. Ignatov, A. Semenov, and J. Poelmans. Gaining insight in social networks with biclustering and triclustering. *LNBIP*, 128:162–171, 2012.
8. E. A. Leicht, P. Holme, and M. E. Newman. Vertex similarity in networks. *Physical Review E*, 73(2):026120, 2006.
9. R. G. Pensa, D. Ienco, and R. Meo. Hierarchical co-clustering: off-line and incremental approaches. *Data Mining and Knowledge Discovery*, 28(1):31–64, 2014.
10. F. Shi, J. G. Foster, and J. A. Evans. Weaving the fabric of science: Dynamic network models of science's unfolding structure. *Social Networks*, 43:73–85, 2015.
11. A. Tanay, R. Sharan, and R. Shamir. Biclustering algorithms: A survey. *Handbook of computational molecular biology*, 9(1-20):122–124, 2005.
12. L. Teng and K. Tan. Finding combinatorial histone code by semi-supervised biclustering. *BMC Genomics*, 13(1):301, 2012.