UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

A Bayesian Metric for Network Similarity

Permalink

https://escholarship.org/uc/item/2r38g0nx

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 38(0)

Author

Shiffrin, Richard M.

Publication Date

2016

Peer reviewed

A Bayesian Metric for Network Similarity

Richard M. Shiffrin (shiffrin@indiana.edu)
Psychological and Brain Sciences / Cognitive Science Program
1101 E 10th Street, Bloomington IN 47405 USA

Keywords: Network Similarity, Network Inference, Network Comparison, Bayesian Methods

Abstract

Networks of every kind and in numerous fields are omnipresent in today's society (e.g. brain networks, social networks) and are the intense subject of research. It would be of great utility to have a computationally efficient and generally applicable method for assessing similarity of networks. The field (going back to the 1950s) has not come up with such a method (albeit a few moves in this direction exist, such as Jaccard coefficients, QAP--quadratic assignment procedure, and more recently Menezes & Roth, 2013, and Asta & Shalizi, 2014). I present a Bayesian-based metric for assessing similarity of two networks, possibly of different size, that include nodes and links between nodes. I assume the nodes are labeled so that both the nodes and links between two nodes that are shared between the two networks can be identified.

The method calculates similarity as (a monotonic transformation of) the odds that the two observed networks, termed V and W, were produced by random sampling from a single master network, termed G, as opposed to generation by two different but similar networks, termed Gv and Gw. The simplest form of the method ignores strengths that could be assigned to nodes and links, and considers only nodes and links that are, or are not, shared by the networks. Suppose there are n_V nodes and N_V links only in V, n_W nodes and N_W links only in W and n_c nodes and N_c links shared between the networks. Thus the number of nodes in V is $n_c + n_V$ and the number in W is $n_c + n_W$. The number of unique nodes in both V and W is $n_c + n_V + n_W = n$. The number of links in V is $N_c + N_V$ and the number in W is $N_c + N_W$. The number of unique links in both V and W is $N_c + N_V + N_W = N$.

The single master network, G, is assumed to consist of the union of the nodes and links in the two networks, and has n nodes and N links. The probability a given shared node will be randomly and independently sampled twice is $[(n_V + n_c)/n][(n_W + n_c)/n]$. The probability a given shared link will be randomly and independently sampled twice is $[(N_V + N_c)/N][(N_W + N_c)/N]$.

If there are two generating networks I assume they each have n nodes and N links. I also assume they are similar,

because we would not be comparing dissimilar networks. The degree of similarity is controlled by 'tuning' parameters[1]: Gv and Gw are assumed to share αn nodes and βN links. The probability a given shared node will be sampled twice is then $\alpha[(n_V + n_c)/n][(n_W + n_c)/n]$, and the probability a given shared link will be sampled twice is $\beta[(N_V + N_c)/N][(N_W + N_c)/N]$. The likelihood ratio λ_{js} for G vs (GV, GW) as generator of a given shared node is then $1/\alpha$ and the likelihood ratio π_{js} of a given shared link is then $1/\beta$.

For a non-shared node, say in V, similar reasoning gives a likelihood ratio λ_{kV} of

$[1-(n_W + n_c)/n)] / [1 - \alpha(n_W + n_c)/n]$

and for a non-shared link a likelihood ratio π_{kV} of

$[1-(N_W + N_c)/n)] / [1 - \alpha(N_W + N_c)/N]$

For a non-shared node or link in W substitute a W subscript for the V subscript in these likelihood ratios.

Computational efficiency is a necessity if the similarity metric is to be applied to large networks. For this reason I do not calculate the exact probabilities for the numbers of shared and non-shared nodes and links that are observed (the combinatoric complexity of such calculations is enormous). Instead I make the simplifying assumption that each node and link contribute the likelihood ratios given above and that the total odds is obtained by multiplying all the likelihood ratios together. This simplification can perhaps be justified if similar distortion is produced by this simplifying assumption for both the cases of G and (G_V, G_W) as generators. Under this simplifying assumption the overall odds becomes:

$$\varphi(1/2) = (\lambda_{js})^{nc}(\lambda_{kV})^{nV}(\lambda_{jW})^{nW}(\pi_{js})^{Nc}(\pi_{kV})^{NV}(\pi_{jW})^{NW}$$

Taking the log of this product converts the calculation to sums and makes calculation highly efficient.

This abstract is too short to permit giving the different and more complex results that hold for the several cases when the nodes and/or links have associated strengths. I give a summary of some of the results here. The results for links and nodes are similar so consider the results for nodes. Let there be just one set of strength values, Si for the i-th node. Norm these to sum to 1.0. For either generation by G or (Gv,Gw) assume sampling is made without replacement and proportional to strength. Let Ziv and Ziw be the probabilities that node i will be sampled by $n_v + n_c$ samples, or $n_w + n_c$ samples respectively. The Z's would be difficult to

obtain analytically but could be estimated by Monte Carlo sampling. Consider two possibilities for the way that Gv and Gw overlap. In Case A the probability a node will be shared is simply α, independent of strength. In Case B, the probability a node will be shared is an increasing function of strength, Y_i.

For Case A the likelihood ratio for a shared node i is: $1/\alpha$. For a node k only in V the likelihood ratio is: $\lambda_{kV} = (1-Zkw)/\{1 - \alpha (1-Zkw)\}$. For a node only in W exchange the subscripts v and w. Then we have for the odds due to nodes: $\varphi_D = (1/\alpha)^{nc}\Pi_k(\lambda_{kV})\Pi_j(\lambda_{jw})$.

For Case B the likelihood ratio for a shared node i is $1/Y_i$. For a node k only in V the likelihood ratio is: $\lambda_{kV} = (1-Zkw)/\{1-Y_k(1-Zkw)\}$. Again switch v and w subscripts for a node only in W. Then we have for the odds due to nodes: $\varphi_D = \Pi_i(1/Y_i)\Pi_k(\lambda_{kV})\Pi_j(\lambda_{jw})$.

These expressions would have analogous forms for links, with different Ns, Z's and Y's, and the overall odds would, as before, be a product of the odds for nodes and the odds for links.

The critical difference between Cases A and B is the degree to which evidence based on an observed shared node or link is strength dependent: For Case B this evidence rises as strength decreases. This should raise concerns: However strengths are obtained there is likely to be measurement noise that reduces the reliability of low strength values. This might argue in favor of using Case A, or if one preferred Case B to restrict the Yi values to lie above a lower bound. The idea would be to let evidence depend most on the nodes (or links with high strength values.

It should be observed that the existence of a computationally efficient and generally applicable metric for network similarity would allow alignment of non-labeled networks. One would search for the alignment of nodes that would maximize the metric.

I have many relevant publications demonstrating some degree of expertise in Bayesian modeling (e.g.: Shiffrin & Chandramouli, in press; Shiffrin, Chandramouli, & Grünwald, 2015; Chandramouli & Shiffrin, 2015; Nelson & Shiffrin, 2013; Cox & Shiffrin, 2012; Shiffrin, Lee, Kim, & Wagenmakers, 2008; Cohen, Shiffrin, Gold, Ross, & Ross, 2007; Denton & Shiffrin; Huber, Shiffrin, Lyle, & Ruys, 2001; Shiffrin & Steyvers, 1997). I note that the present results are in a vague sense an extension of the metric proposed for matching memory probes to memory traces that are given in Cox and Shiffrin (2012) and in the appendix of Nelsonb and Shiffrin (2013).

References

Asta, D., & Shalizi, C. R. (2014). Geometric Network Comparison. arXiv preprint arXiv:1411.1350.

Busemeyer, J. R., Zheng, W., & Shiffrin, R. M. (2015). Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency. *Decision*, 2, 1-12.

Chandramouli, S. H., & Shiffrin, R. M. (2015). Extending Bayesian induction. *Journal of Mathematical Psychology*. http://dx.doi.org/10.1016/j.jmp.2015.10.002

Cohen, A. L., Shiffrin, R. M., Gold, J. M., Ross, D. A., & Ross, M. G. (2007). Inducing features from visual noise. *Journal of Vision*, 7(8):15, 1-14.

Cox, G. E., & Shiffrin, R. M. (2012). Criterion setting and the dynamics of recognition memory. *Topics in Cognitive Science*, 4(1), 135-150.

Denton, S. E., & Shiffrin, R. M. (2012). Primes and flankers: Source confusions and discounting. *Attention, Perception, & Psychophysics*, 74 (5), 852-856.

Huber, D. E., Shiffrin, R. M., Lyle, K. B., & Ruys, K. I. (2001). Perception and preference in short-term word priming. *Psychological Review*, 108, 149-182.

Menezes, T., & Roth, C. (2014). Symbolic regression of generative network models. *Scientific reports*, 4.

Nelson, A. B. and Shiffrin, R.M. (2013). The co-evolution of knowledge and event memory. *Psychological Review*, 120(2); 356-394.

Shiffrin, R. M., & Chandramouli, S. H. (in press). Model selection, data distributions, and reproducibility. In H. Atmanspacher and S. Maasen (Eds.) *Reproducibility: Principles, Problems, and Practices*. John Wiley, New York.

Shiffrin, R. M., Chandramouli, S. H., & Grünwald, P. D. (2015). Bayes Factors, relations to Minimum Description Length, and overlapping model classes. *Journal of Mathematical Psychology*. http://dx.doi.org/10.1016/j.jmp.2015.11.002

Shiffrin, R. M., Lee, M. D., Kim, W. & Wagenmakers,, E. J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, 32, 1248-1284.

Shiffrin, R. M., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review*, 4(2), 145-166.

1. α and β would have to be quite large to produce reasonable similarity values. Research will be needed to determine whether the values can be fixed for all types of networks being compared, or adjusted for different network types.