Inferring monopartite projections of bipartite networks: an entropy-based approach

F. Saracco[1], M. Straka[1], R. Di Clemente[2], A. Gabrielli[3], G. Caldarelli[1], T. Squartini[1]

[1]IMT School for Advanced Studies Lucca

[2]Massachusetts Institute of Technology

[3]Institute for Complex Systems (ISC-CNR), "Sapienza" University

Bipartite networks are currently regarded as providing a major insight into the organization of many real-world systems, unveiling the mechanisms driving the interactions which occur between distinct groups of nodes. One of the most important issues encountered when modeling bipartite networks is devising a way to obtain a (monopartite) projection on the layer of interest, which preserves the information encoded into the original bipartite structure as much as possible.

In the present paper we propose an algorithm to obtain statistically-validated projections of bipartite networks, which implements a simple rule: in order for any two nodes to be linked, the number of shared neighbors must be statistically significant. Our assumption can be translated into requiring that the number of observed V motifs between the considered pair of nodes is significant.

Naturally, assessing the statistical significance of nodes similarity requires the definition of a proper statistical benchmark: here we consider a set of four null models, defined within the Exponential Random Graph framework. These benchmarks differ by the amount of information constrained (while the Bipartite Random Graph Model is only defined by the total number of links, the Bipartite Configuration Model constrains the whole degree sequence), thus providing benchmarks inducing projections characterized by a different level of detail. As a result, our algorithm outputs a matrix of link specific p-values, from which a validated projection can be straightforwardly obtained, upon running a multiple hypothesis test and retaining only the statistically-significant links.

Finally, we have tested our method on an economic network (i.e. the countries-products World Trade Web representation) and on a social network (i.e. the MovieLens dataset, collecting the users ratings of a list of movies): in both cases non trivial communities are detected.

In the first case, while projecting the World Trade Web on the countries layer reveals modules of similarly-industrialized nations, projecting it on the products layer allows communities characterized by an increasing level of complexity to be detected; in the second case, projecting MovieLens on the films layer allows clusters of movies whose affinity cannot be fully accounted for by genre similarity to be individuated.

[l] Benjamini, Hochberg, J. R. Stat. Soc. B **57**, 289 (1995).

[2] Tumminello et al., PLoS ONE **6**, e17994 (2011).

[3] Neal, Soc. Netw. **39**, 84 (2014).

[4] Latapy et al., Soc. Netw. **30**, 31 (2008).

[5] Gualdi et al., Sc. Rep. **6**, 39467 (2016).