Modularity and edge sampling

The modularity of a graph is a measure of the extent to which the graph breaks into separate communities. For a given graph G, each partition $\mathcal{A}$ of the vertices has a modularity score $q_{\mathcal{A}}(G)$, with higher values indicating that the partition better captures community structure in G. The modularity $q^*(G)$ of the graph G is defined to be the maximum over all vertex partitions of the modularity score, and satisfies $0 \leq q^*(G) \leq 1$. Modularity based community detection algorithms are commonly used in the analysis of network data.

We analyse when the modularity of an underlying graph can be inferred from its observed graph: suppose edges in an underlying graph G appear with some probability in our observed graph G'. If G has high modularity, is the observed graph G' likely to have high modularity? We see that this is indeed the case. We find that $q^*(G') \geq q^*(G) - \varepsilon$ with probability at least $1 - \varepsilon$, as long as the expected number of edges in G' is large enough. Similarly, $q^*(G') \leq q^*(G) + \varepsilon$ with probability at least $1 - \varepsilon$, under the stronger condition that the expected average degree in G' is large enough.

Joint work with Colin McDiarmid