

1 Attribute search

Consider a graph $G_i = (V_i, E_i)$ with vertices V_i representing people at iteration i , and edges E_i representing friendships between people at iteration i . Two vertices are neighbors if they share an edge. According to Dunbar's number, a typical person can maintain 150 meaningful relationships with people in their social network; therefore a vertex in our network can have at most 150 neighbors.

We have a set of fixed length vectors A , where A_v represents the vector of attributes of vertex $v \in V_i$. Attributes are chosen from a fixed large number of attributes according to a Pareto distribution. Therefore, we expect most attributes to be rare (shared by few vertices). We refer to the attributes in A_v as the contexts of v . Note that A may change at each iteration.

We additionally define two fixed costs - direct and indirect. Each is motivated by the cost of maintaining a friendship. A person incurs a direct cost to maintain each of their friendships, and an indirect cost to monitor any of their friends' relationships with mutual friends. In our network, a vertex v incurs a fixed direct cost c_d for each of its neighbors, and a fixed indirect cost c_i for each of its neighbors that has an edge with another of v 's neighbors. Each vertex's budget is fixed, and defined by the maximum number of neighbors times the fixed cost for a direct edge, $150 \cdot c_d$.

At iteration 0, to determine A_v for each $v \in V_0$, we assign attributes as described above. For $i > 0$, each context in A_v will switch to inactive with a fixed probability p . If a context in A_v switches to inactive, it is replaced with an attribute chosen as described for iteration 0. Only active contexts (in A_v at iteration i) are used as a basis for forming edges at iteration i .

At each iteration we consider each vertex $v \in V_i$ and use a simple sigmoid function to determine whether an edge can exist between v and $u \in V_i$. The function determines edge formation while accounting for rarity of shared contexts. If the edge can exist, it is added to a list of edge candidates, which is shuffled before the edges are added to the graph. An edge can be added only if the cost of adding the it for each vertex is less than each of their remaining budgets. This new set of edges is E_i . Note that edges are recalculated each iteration, so those that existed at time i may not at time $i + 1$.

After network formation, our goal is to calculate the probability of edge formation as a function of distance. Distance is defined as the length of the shortest path from one vertex to another, excluding the direct edge. We compute these probabilities to show that shortcuts in the graph are in fact post-hoc relationships - edges form based on shared contexts, and when vertices lose these contexts, the edges remain.

Add formal description