Social and Technological Networks	Edinburg, 2015

Lecture 5 & 6. Small Worlds

Rik Sarkar *Class notes*

We covered the following in class:

- Milgrams experiment
- A definition of small world networks
- The Watts Strogatz model for small worlds
- Kleinberg's model for small worlds

Note that different people tend to have different things in mind when they say *small world*. The definition we had in class is one I thought reasonable. At the very least, a small world should have small diameter and high clustering coefficient, preferably also decentralized search. What is definitely not correct is to say that a graph with small diameter is a small world graph. Although this is a mistake a lot of people make, including sometimes in published papers.

1 Watts-Strogatz model

Note that in the original paper, the authors suggest using $k > logn$ neighbors per node, but very often, k is taken to be a constant, because people will usually have limited friend-circles. We will stick to this idea for our discussions and assume k is a constant.

1.1 Clustering coefficient per vertex

We want to show that the clustering coefficient at any vertex is bounded below by a constant.

Exercise 1.1. *If the number of neighbors k is a constant in a Watts-Strogatz graph, (whereas n is not a constant – the graph can be very large), and the fraction of rewired edges β is also a constant, show that there exists a constant c such that for any vertex u, the expected clustering coefficient is grater than c.*

1.2 Not finding short paths in Watts-Strogatz graphs

We used a slightly simplified model in class, where the *long links* are added onto the underlying $n \times n$ grid instead of the grid being modified. The results hold in the original version too, just that this one is a little easier to discuss.

We showed in class that in this model, it is hard to find short paths. The idea goes as follows. Suppose s and t are distance $n^{2/3}$ apart. Then our analysis (see slides and [Kempe, 2011]) shows that it is unlikely that a long link from s will take us to a node nearer to t. Thus, a local routing or search algorithm must take a short step (1 unit length or any constant) from s and look again. Once again, it is unlikely to find a helpful long link, and continue with a short hop. This will continue until it finds a helpful long link.

We show that the expected number of such short hops before finding a useful long link is $\approx n^{2/3}$, but since that was the distance (using short hops) anyway, the algorithm in effect never found a useful long link.

* **Exercise 1.2.** *Suppose we start with s and t at distance $n^{1/4}$. Does the above situation hold, that the local routing algorithm will in expctation have to go the entire way without a useful long link? [hint: what is the probability that a long link lands within distance $n^{1/4}$ of t, and therefore how many hops will it require to find a useful long link?]*

Try the same with a higher distance, like $n^{3/4}$. Why do you think we used $n^{2/3}$ for the analysis in class?

1.3 Kleinberg's model

We saw that Kleinberg's model fixes the decentralized search. The problem with Watts-Strogatz model was that once the search got kind of close to the destination (still a polynomial distance at $n^{2/3}$), it was unlikely to find a long link that will land in the region closer to the destination. A link that was in the right direction was still likely to overshoot the destination, simply because there are more nodes far from the destination than those close to the destination. Kleinberg's model fixes this by increasing probabilities of shorter long links so that as we get closer to destination, they are unlikely to overshoot. But it still keeps enough long-long links that help in quickly getting closer when starting from a distance.

Exercise 1.3. *Suppose instead of a $2D$ grid, we started with a $3D$ grid. What do you think is the distribution needed for decentralized search to work with the same $log^2 n$ guarantee?*

1.4 A few additional concepts: Metric, Growth and doubling dimension

Growth. The number vertices of balls of different sizes is usually referred to as a quantity called the *growth*. Growth is usually measured in the hop-distance metric of the graph. To say that the growth at a vertex v is a function g means that $|B_r(v)| = g(r)$. Usually the growth is considered in asymptotic notation:

- At any vertex on a $2D$ grid, growth is $\Theta(r^2)$ (polynomial)
- At any vertex on a balanced binary tree, growth is $\Theta(2^r)$ (exponential)

To say that a graph has a certain asymptotic growth usually implies that the growth at any vertex satisfies that growth rate.

Note that depending on the graph, or what we know about it, the asymptotic growth may not always satisfy a $\Theta(g(r))$ form, but may have a bound in only one direction such as $O(r^2)$ or $\Omega(r^3)$ etc.

Growth is a useful notion in making arguments that are based on number of vertices in a ball centered at a vertex as we were using in this topic.

Exercise 1.4. *Show that graphs with constant expansion have exponential growth.*

Spheres. A sphere $S_r(v)$ is the set of vertices exactly at a distance r from v.

Doubling dimension. Suppose $\mathcal{S}$ is a set of subsets of V. Then we say that $\mathcal{S}$ *covers* V if every $v \in V$ belongs to some $s \in S$.

Suppose $\mathcal{S}$ is a set of balls in the graph, then we say $\mathcal{S}$ covers the graph if every vertex of the graph is in some ball in $\mathcal{S}$. You can imagine a grid in the plane, and the balls as disks that *cover* the area of the grid. (this is not exactly precise, but good as a starting point).

Definition 1.1 (Doubling dimension). *A graph is said to have a doubling dimension η if for every vertex v, the ball of radius r, that is $B_r(v)$, can be covered using at most 2^η balls of radius $r/2$, for all r.*

This is an adaptation of the notion of dimension to discrete structures such as graphs. Our main interest is whether a graph has its doubling dimension bounded by a constant or not. The grids ($2D, 3D, \dots$) have constant doubling dimension (can you find the constants?)

Exercise 1.5. *Show that the balanced binary tree does not have a bounded doubling dimension.*

Metrics. A metric is a measure of distances that satisfies triangle inequality. See wikipedia for an exact definition.

The hop-distance in a graph satisfies triangle inequality and is a metric (we define the distance between two vertices as the shortest path distance in the graph). Our usual familiar distances in the plane or space are also metrics. As are distances on a piece of bent metal or crumpled piece of paper. Though in these cases, there are no vertices, but a continuum of points.

All the notions we saw above also applicable to metric spaces. Except that Since the number of points is uncountable, we cannot count them like vertices. Instead we use the notion of *volume* in suitable dimensions. That is, we consider area as the volume in $2D$, length as volume of a $1D$ object etc.

You can check that the $2D$ euclidean plane has a bounded doubling dimension.

References

[Kempe, 2011] Kempe, D. (2011). Structure and dynamics of information in networks, lecture notes. Technical report, U.S.C.