Review of Literature

The Small World Phenomenon: An Algorithmic Perspective

Jon Kleinberg

Reviewed by:

Siddharth Srinivasan

Oh, it’s such a small world!!

- Milgram (1967, 69) – performed an empirical validation of the small world concept in sociology.
  - Previous work-
    - Pool and Kochen model 2 people at random connected with k intermediaries. Assumes synthetic, homogenous structure.
    - Rapaport and Horvath – empirical study on school friendships. Asymmetric nets and Universe is small.
- Packet sent by a randomly chosen source to a random target.
  - Mean chain length = 5.2
  - Variables of geographic proximity, profession and sex
  - Funneling of chains by certain individuals

Small world! Small world!

- White (1970) – tries fitting a simple model to Milgram's work.
  - Gives hints to future work
- Killworth & Bernard (1979) – Reverse SW
  - To understand social network structure, factors that influence the choice of acquaintance, the out-degree of people.
  - Results:
    - Generation of contacts not purely random.
    - Large number of contacts for local targets; few contacts for non-local targets.
    - The size of geographical area that a single contact is responsible for decreases as a function of the distance of the target from starter.
    - Most choices based on cues of occupation and geographic location.

Small Worlds Everywhere

- Watts and Strogatz (1998)
  - Very small number of long range contacts needed to decrease path lengths without much reduction in cliquishness.
  - Long range contact picked uniformly at random (u.a.r)
  - Small world networks in 3 different areas esp. spread of infectious disease.
    - Probabilistic reach. No specific destinations.
    - Doesn’t require knowledge of paths and no active path selection.
- Barabasi et al.(1999) – diameter of the WWW
  - Power-law distribution; Logarithmic diameter.
  - Need for search engines to intelligently pick links

Two Important Properties of Small World Networks

- Low average hop count
- High clustering coefficient

Additionally, may be searchable on the basis of local information

Enter Kleinberg…

- Two issues of concern in small-world networks:
  - **<u>Presence of short paths in a small world network</u>**
  - **<u>how do you find the short chains?</u>**
- Gives an infinite family of small world network models on a grid n/w with power-law distributed random long-range links.
  - K(n,k,p,q,r)
  - p – radius of neighbours to which short, local links
  - q – no. of random long range links
  - k - dimension of mesh (k=2 in this paper)
  - r - clustering exponent of inverse power-law distribution.
  - Prob.[(x,y)] $\propto$ dist(x,y)$^{-r}$.
- Decentralized greedy routing algorithm
  - Decisions based on local information only.

Bounds on Kleinberg's Model

- Expected Delivery time =
  - $O((\log n)^2)$, for r = 2.
  - $\Omega(n^{(2-r)/3})$, for $0 \le r < 2$.
  - $\Omega(n^{(r-2)/(r-1)})$, for $2 < r$
- Disproves usefulness of Watts & Strogatz model (r=0).
- Only for special case of r = k, possible to find short chains always of length $O((\log n)^2)$ and dia = $O(\log n)$.
- Cues used in small world networks propounded to be provided through a correlation between structure and distribution of long-range connections.

Proof of the upper bound

- For r=2, p=1, q=1.
- Event $E_u(v)$ - u chooses v as its random long range contact
- Prob[$E_u(v)$] = $d(u,v)^{-2} / \sum_{v \neq u} d(u,v)^{-2}$
- $\sum_{v \neq u} d(u,v)^{-2} \leq \sum_{j=1}^{2n-2} (4j)(j^{-2}) \leq 4 + 4\ln(2n-2) \leq 4\ln(6n).$
- $\therefore$Prob[$E_u(v)$] ≤ $[4 \ln(6n)\, d(u,v)^2]^{-1}$.
- In phase j, $2^j < d(u,t) \leq 2^{j+1}$. For log(log n) < j < log n,
  - No. of nodes in Bj ≥ $1 + \sum_{i=1}^{2^j} i = \frac{1}{2}2^{2j} + \frac{1}{2}2^j + 1 > 2^{2j-1}$
    each within lattice distance $2^j + 2^{j+1} < 2^{j+2}$ of u
  - Prob[Enters Bj] ≥ $\frac{2^{2j-1}}{4\ln(6n)2^{2j+4}} = \frac{1}{128\ln(6n)}$.
  - Steps in j = X_j;
    $EX_j = \sum_{i=1}^{\infty} \Pr[X_j \geq i] \leq \sum_{i=1}^{\infty}\left(1 - \frac{1}{128\ln(6n)}\right)^{i-1} = 128\ln(6n).$
- $\therefore$ $EX \leq (1 + \log n)(128\ln(6n)) \leq \alpha_2(\log n)^2$

Proof of lower bound 1

- As in the previous proof,

$$\sum_{v \neq u} d(u,v)^{-r} \;\geq\; \sum_{j=1}^{n/2} (j)(j^{-r}) \;\geq\; \int_1^{n/2} x^{1-r}\, dx \;\geq\; \frac{1}{(2-r)2^{3-r}} \cdot n^{2-r}$$

where, assumed that $n^{2-r} \geq 2^{3-r}$.

- Let $\delta = (2-r)/3$ and U be the set of nodes witihin radius pn^{δ} of t.

$$|U| \leq 1 + \sum_{j=1}^{pn^{\delta}} 4j \leq 4p^2 n^{2\delta}$$ where, assumed that $pn^{\delta} \geq 2$.

- Let ε' be the event that the msg reaches a node in $U \neq t$ in λn^{δ} steps. Let ε'_i be the event that this happens in the i^{th} step.

$$\Pr[\mathcal{E}'_i] \;\leq\; \frac{q|U|}{\frac{1}{(2-r)2^{3-r}} \cdot n^{2-r}} \;\leq\; \frac{(2-r)2^{3-r} q \cdot 4p^2 n^{2\delta}}{n^{2-r}} \;=\; \frac{(2-r)2^{5-r} qp^2 n^{2\delta}}{n^{2-r}}$$

- $$\Pr[\mathcal{E}'] \;\leq\; \sum_{i \leq \lambda n^{\delta}} \Pr[\mathcal{E}'_i] \;\leq\; \frac{(2-r)2^{5-r} \lambda qp^2 n^{3\delta}}{n^{2-r}} \;=\; (2-r)2^{5-r}\lambda qp^2 \leq \frac{1}{4}.$$

  where $\lambda = (2^{8-r} qp^2)^{-1}$

Proof of lower bound 1 contd.

- Let events F (s and t separated by ≥ n/4).

  $Pr[F] \geq ½$; $Pr[!F \vee \varepsilon'] \leq ¾$; and so $Pr[F \wedge !\varepsilon'] \geq ¼$.
- Let ε - event that msg reaches t from s in λn^{δ} steps.

  ε cannot occur if $(F \wedge !\varepsilon')$ occurs.
- $\therefore Pr[\varepsilon \mid (F \wedge !\varepsilon')] = 0$ and $E[X|(F \wedge !\varepsilon')] \geq \lambda n^{\delta}$ steps.
- $E[X] \geq E[X|(F \wedge !\varepsilon')] \,.\, Pr[F \wedge !\varepsilon'] \geq ¼\lambda n^{\delta}$ steps,

  where, X is the random variable denoting the no. of steps.
- Thus, lower bound on <u>expected no. of steps is $\Omega(n^{(2-r)/3})$, for $0 \leq r < 2$</u>.

Proof of lower bound 2

- Similar to the previous proof, $$\Pr[d(u,v) > m] \leq \sum_{j=m+1}^{2n-2} (4j)(j^{-r}) \leq (r-2)^{-1} m^{2-r} = \varepsilon^{-1} m^{-\varepsilon}.$$

  where, ε = r-2.
- Let β = ε ⁄ 1+ε, γ = 1 ⁄ 1+ε, and λ’ = min(1,ε)/8q. Assumed that $n^{\gamma} \geq p$.
- Let ε'_i be the event that in the i^{th} step, msg reaches u w/ a long range contact v such that $d(u,v) > n^{\gamma}$.

  Let ε’ be the event that this happens in $\lambda' n^{\beta}$ steps.

  $$\Pr[\mathcal{E}'] \leq \sum_{i \leq \lambda' n^{\beta}} \Pr[\mathcal{E}_i] \leq \lambda' n^{\beta} \cdot q\varepsilon^{-1} n^{-\varepsilon\gamma} = \lambda' q \varepsilon^{-1} \leq \frac{1}{4}.$$
- Similar to the previous proof,
- max dist. Covered w/o ε’ occuring is $\lambda' n^{\beta+\gamma} = \lambda' n < n/4$ and hence,

  $$EX \geq E\left[X \mid \mathcal{F} \wedge \overline{\mathcal{E}'}\right] \cdot \Pr\left[\mathcal{F} \wedge \overline{\mathcal{E}'}\right] \geq \frac{1}{4} \lambda' n^{\beta};$$
- Thus, lower bound on <u>expected no. of steps is $\Omega(n^{(r-2)/(r-1)})$, for 2 < r</u>

Major Ideas Contributed

- Gives a model of a small world network where local routing is possible using small paths.
- Shows the more generalized results for k dimensions in a subsequent publication.
- Correlation between local structure and long range links provides fundamental cues for finding paths.
  - When r<k, few cues provided by the structure
  - When r>k, long range links do not provide sufficiently long jumps and path becomes long.

Questions Raised

- Can the expected delivery time be reduced to the bounds of the diameter?
- Is the model extendable to more general networks?
- Can less regular base graphs also produce navigable small worlds?

Work Done post-papyri

- Further analysis and generalization of Kleinberg’s models and other small world models
- Conversion of general networks to small world networks
- Applications of the small world idea to real networks

Further Analysis and Generalizations 1

- Barriere et al.(2001) –
  - proves $\Theta((\log n)^2)$ bound on routing complexity. Simplified analysis using a ring instead of a grid.
  - Oblivious greedy routing.
  - Basic concept used in analysis – (f, c)-long range contact graph – if for any pair (u,t) at distance at most d, we have $\Pr[u \rightarrow B_{d/c}(t)] \geq 1/f(d)$.
  - If graph (G, p) is an (f, c)-long range contact graph then greedy routing in $O(\sum_{i=1}^{\log cD} f(D/c_i))$ expected steps.
  - If p is a non-decreasing fn., then $\Pr[u \rightarrow B_{d/c}(t)] \geq \Pr[(c+1)d/c] \,.\, |B_{d/c}(t)|$
  - extends results to any ring by epimorphisms (embedding) one graph to another.

Further Analysis and Generalizations 2

- Martel, C. and Nguyen, V. (2004):
  - Shows that Kleinberg's algo is tight $\Theta(\log^2 n)$ expected delivery time and diameter tight at $\Theta(\log n)$.
  - For k-dimensional grid as well.
  - If additional info, then $O(\log^{3/2} n)$ for k=2 and $O(\log^{1+1/k} n)$ for $k \geq 1$.
  - Proof done in a manner that uses some interesting conceptual ideas (used by others previously as well):
    - $p(u, v) = d^{-2}(u, v)/c_u$, $c_u = \sum d^{-2}(u, v) = \sum b_j(u)\, j^{-2}$;
    - $b_j(u) = \Theta(j)$, so, c_u approx. as a harmonic sum.
    - Inherently uses the concept of gradient, $\delta(v) = d(v,t) - d(N(v),t)$, to show the lower bound.
    - Uses the concept of harmonics to get for any integer $1 < m < d(v, t)$:

$$\Pr[\delta(v) = m] \leq \frac{\hat{c}}{m \log n} \times \min\{1, \frac{d(v,t)-m}{m}\}$$

    - Expected delivery time is $\Omega(\log^2 n)$ for any s and t w/ probability ≥ 0.5 when $d(s,t)$ is $O(n)$.

- Extended algo – Window (no. of neighbouring nodes whose long range contacts are known) = log n.
  - In k dimensions, $O(\log^{1+1/k} n)$. Prove only for k=2.
- Diameter = $\Theta(\log n)$. Extended to all possible K|K*(k,n,p,q) where k, p, q ≥ 1 and even for 0<r<2.
  - grow trees from s and t using only long-range links starting from an initial set of size $\Theta(\log n)$ and going upto a set of size $\Theta(n\log n)$ in $O(\log n)$ steps. With very high probability, these sets will overlap or be separated by a single link.
- Extensions based on concept of developing supernodes (composite of neighbouring nodes to get all their random links) for analysis.
- Subsequent work shows that
  - poly-log expected dia. when k<r<2k
  - Polynomial expected dia. when r>2k.

Further Analysis and Generalizations 3

- Fraigniaud et al. (2004) – “Eclecticism shrinks even small worlds”
  - Dimensions need not mean only geographical dimensions but can refer to the various parameters used for routing in social networks – geography, occupation, education, socio-economic status etc.
  - Higher dimensions intuitively must give better performance,
    - dimension not considered in routing performance in the greedy algo proposed by Kleinberg since $O(\log^2 n)$ in all dimensions.
  - Giving $O(\log^2 n)$ bits of topological awareness per node decreases the expected number of steps of greedy routing to $O(\log^{1+1/k} n)$ in k-dimensional augmented meshes.

- Called indirect greedy routing. Completely oblivious routing.
- Analysis proves that between two nodes in a sequence of long-range nodes, $dist(z_i, z_{i+1}) \leq \log^{1/k} n$. And, totally $O(\log n)$ such nodes.
- Augmenting the topological awareness above this optimum of $O(\log^2 n)$ bits would drastically decrease the performance of greedy routing.
- Perhaps a first step towards the formalization of arguments in favor of the sociological evidence stating that eclecticism shrinks the world.

Further Analysis and Generalizations 4

- Raghavan et al. (2005). “Theoretical Analysis of Geographic Routing in Social Networks.”
  - rank-based friendship - probability that a person v is a friend of a person u is inversely proportional to the number of people w who live closer to u than v does.
    - $rank_u(v)$ = no. of people w such that $d(u,w) < d(u,v)$.
  - $prob(u,v) = rank_u(v)^{-1}$.
  - more accurately models the behaviour of social networks – verified against LiveJournal data.
  - in a grid setting, $prob(u,v) = rank^{-1} = d^{-k}$.
  - Halves distance in expected polylogarithmic steps –
    - Starting from s, expected number of steps before reaches a point in $B_{d(s,t)/2}(t)$ is $O(\log n \log m) = O(\log^2 n)$
  - Finds short paths in all 2-D meshes –
    - For any 2-dimensional mesh population network with n people and m locations, expected path length is $O(\log n \log^2 m) = O(\log^3 n)$.
  - Interesting proof methodology – using only balls. Plus rank and balls is general over all dimensions.

Further Analysis and Generalizations 5

- Watts et al. (2002) and Motter et al. (2003).
  - hierarchies of social groups with groups having some correlation between them.
  - social ties generated by picking links from social groups according to p.distribution governed by social affinity.
- Manku et al. (2004). Know thy neighbour's neighbour.
  - Shows that if every node is aware of the long-range links of its neighbours then greedy routing in $O(\log^2 n/(c \log c))$ with c long range contacts per node.

Conversion to small world networks

- Duchon et al. (2006). At INRIA
  - On bounded growth graphs and extended to polylogarithmic expansion rates.
  - Using O(n) rounds and O(polylog n) space. No need for a node to have complete knowledge of the graph.
  - Any synchronized n-node network of bounded growth, of diameter D, and maximum degree Δ, can be turned into a small world via the addition of one link per node,
    - in O(n) rounds, with an expected number of messages O(nD log n), and requiring O(Δ log n logD) memory size with high probability, or,
    - in O(D) rounds with an expected number of messages O(nlog D log n), and requiring O(n) bits of memory in each node with high probability
  - In the augmented network, the greedy routing algorithm computes paths of expected length O(logDlog δ + log n) between any pair of nodes at mutual distance δ in the original network.
  - Sampling of leader nodes.
    - Only leader nodes explore a ball $B_v(3l)$, when asked by a node u at a distance ≤ l ($l=2^i$), to select a random long range link for it, where i is selected u.a.r.

Some Applications Areas

- P2P overlay networks
- Distributed hashing protocols
- Security systems in mobile ad hoc networks
- Hybrid sensor networks
- Referral systems

Applications: Distributed Hashing

- Manku et al. (2002) – Symphony
  - arrange all participants in a ring I [0,1).
  - A node manages that sub-range of I which corresponds to the segment between itself and its two neighbours
  - equip them with long range contacts
    - drawn randomly from a family of harmonic distributions
    - $p = 1/(x \ln n)$ where $x \in [1/n, 1]$ drawn u.a.r.
  - advantages – low degree, can handle heterogeneity by variable number of long range links and only two mandatory short links, low latency $O((\log n)/k)$.
  - for fault tolerance, add f number of backups but only on the short link neighbours.

Applications: P2P Overlay Networks

- Bonsma (2002) - SWAN (Small World Adaptive Network)
  - each node has 3 types of links – bootstrap, local (short-range) and long-range (random).
- Hui et al. - SWOP (Small World Overlay Protocol)
  - Cluster links and long links
  - Head nodes and inner nodes
  - Pdf: Prob[X'=x] = p(x) = 1/(x ln m) where, $x \in [1,m]$ and m is no. of clusters
  - To handle flash crowds, demand-driven replication over long links.

Applications: Hybrid Sensor Networks

- Sharma & Mazumdar (2005) –
  - Adding of a few shortcut wires between wireless sensors.
  - Reduced energy dissipation per node as well as non-uniformity in expenditure.
  - Deterministic as well as probabilistic placement of wires.
  - Few wires unlike 1 long range contact per node in Kleinberg's model. One in a cell / group of cells of sensors is wired.
  - Very good performance in static sink node case
    - with addition of $\Theta(nl(n)/\log n)$ wires, average hop count reduced to $\Theta(1/\sqrt{l(n)})$ and EDS to $\Theta(1/l(n))$.
  - In dynamic case, with greedy routing, hop count cant be reduced below $\Omega(1/l(n))$.

Applications: Security Systems in Ad Hoc N/ws

- Hubaux et al. (2002).
- Gray et al. (2003). Trust propagation

Bibliography

1. Albert, Jeong, Barabasi (1999). Diameter of the World Wide Web, Nature.
2. Barriere, Fraigniaud, Kranakis, Krizanc (2001). Efficient routing in networks with long range contacts
3. Bonsma and Hoile (2002). A distributed implementation of the SWAN peer-to-peer look-up system using mobile agents.
4. Duchon, Hanusse, Lebhar, Schabanel (2006). Fully distributed scheme to turn a network in to a small world. Research report No. 2006-03, INRIA Lyon.
5. Fraigniaud, Gavoille, Paul (2004). Eclecticism shrinks even small worlds.
6. Gray, Seigneur, Chen, Jensen (2003). Trust propagation in small world networks.
7. Helmy, A. (2003). Small Worlds in Wireless Networks. IEEE Commun. Lett., vol.7, no.10, pp. 490-492, Oct. 2003. G/A, 14.
8. Hawick & James (2004). Small-World Effects in Wireless Agent Sensor Networks.
9. Hubaux, J.P., Capkun, S., Buttyan, L., (2002). Small Worlds in Security Systems: an Analysis of the PGP Certificate Graph. In: New Security Paradigms Workshop, Norfolk, VA.
10. Hui, Lui, Yau (2006). Small world overlay P2P networks: construction and handling dynamic flash crowds. Accepted in J. of Comp. Networks.
11. Killworth, Bernard (1979). Reverse Small World Experiment, Social Networks.
12. Kleinberg (2000). Navigation in a small world, Nature.

14. Manku, Bawa, Raghavan (2003). Symphony: Distributed hashing in a small world. USENIX Symposium on Internet Technologies and Systems.
15. G. Manku, M. Naor, and U. Wieder (2004). Know Thy Neighbor's Neighbor: The Power of Lookahead in Randomized P2P Networks. In 36th ACM Symp. On Theory of Computing (STOC).
16. Martel, C. and Nguyen, V. (2004). – Analyzing Kleinberg's (and other) small world networks. ACM *PODC '04.*
17. Milgram, Travers (1969). An experimental study of the small world problem, Sociometry.
18. Motter, Nishikawa and Lai (2003). Large scale structural organization of social networks. Physical Review.
19. Raghavan, Kumar, Liben-Nowell, Novak, Andrew Tomkins (2005). Geographic Routing in Social Networks.
20. Raghavan, Kumar, Liben-Nowell, Novak, Andrew Tomkins (2005). Theoretical Analysis of Geographic Routing in Social Networks.
21. Sharma, Mazumdar (2005). Hybrid Sensor Networks: a small world.
22. Watts and Strogatz (1998). Collective dynamics of small world networks, Nature.
23. Watts, D., Dodds, P., Newman, M.: Identity and Search in Social Networks. Science, 296 (2002) 1302–1305
24. White (1970). Search parameters for the small world problem, Social Forces.
25. Yu, Singh (2003). Searching social networks.