Small Worlds
Notes on a master course given by P. Fraigniaud (Fall 2003) and some more recent results

Nicolas Nisse

MASCOTTE, INRIA, I3S(CNRS,Univ. Nice Sophia Antipolis)

1 Six degrees of separation

Everything starts with the experiment done by Milgram [Mil67]. Around 100 people had to send letters to one person they knew the name, the address (Cambridge, MA) and some other practical informations (job, hobbies, etc.). To send the letter, people were guided by the informations about the destination, but under the constraint that they had to transmit the letter to someone they know (hand to hand). It appears that 20 letters arrived and that chains had length between 2 and 10. The average length of the successful paths was 5 (meaning that the letters passed in hands of 6 people). From there, the idea arises that people on earth are six steps away one from each other. This is the *six degrees of separation.*

In 2003, Dodds, Muhamad and Watts tried another experiment using e-mails [DMW03]. There were around 25000 sources and 12 different destinations. Around 400 chains were successful with a average length of 4 (roughly between 1 and 10). Lot of unsuccessful emails were due to laziness or due to the fact that people did not trust in such an experiment.

From these experiments two main remarks arise: there exist short paths between humans and it is possible to find them. This can be turned into two natural questions:

1. why do there exist short chains between humans?
2. how can we find them?

From these questions, a (one among many) definition of small worlds can be stated as follows

Definition 1 (Small World). *A network has the* small worlds properties *if it has small diameter and that short routes (with (poly)logarithmic length) can be found by a greedy algorithm.*

Other definitions exist, all of them requiring the small diameter. Having a small clustering coefficient (friends of my friends are my friends) is another common required property, but we do not consider it in this note.

2 Augmented Graphs

To answer above questions, Watts and Strogatz propose to model small worlds by augmented graphs [WS98].

Definition 2 (Augmented Graph). *An* augmented graph *consists of a pair* $(G, \mathcal{D})$ *where* $G = (V, E)$ *is a graph and* $\mathcal{D}$ *is a probability distribution defining, for any* $u, v \in V$*, the probability to have an extra arc* (u, v)*.*

The extra arcs (that do not belong to E) are called the *long links*. It is important to note that they are chosen independently. In the model of augmented graphs, nodes represent humans and edges/arcs represent social links. The graph G is known by all nodes and represents the global knowledge (geography, professional informations, etc.). On the other hand, an (oriented) long link (u,v) is only known by u and v and it is supposed to model social links that are "not predictable" (hazard friendship, etc.). Let L be the set of such long links.

No, we describe how the *(decentralized) greedy routing algorithm* performs in an augmented graph. For any $v \in V$, let $N_G(v) = \{u \in V \ : \ \{v,u\} \in E\}$ and $N_{\mathcal{D}}(v) = \{u \in V \ : \ (v,u) \in L\}$.

Definition 3 ((Decentralized) Greedy Routing). *When a node v receives a message with destination $d \in V$, $d \neq v$, then v sends the message to its neighbor $u \in N_G(v) \cup N_{\mathcal{D}}(v)$ such that the distance between u and d in G (without considering the long links) is minimum. Ties are broken uniformly at random.*

The following question has been widely studied during the last decade.

Question 1. Given a graph G, does there exist a probability distribution $\mathcal{D}$ such that the augmented graph $(G,\mathcal{D})$ is a small world?

3 Augmenting a D-dimensional grid [Kle00b]

Let G_D be a D-dimensional grid, $D > 0$, with n vertices. Let $r \geq 0$. We consider the probability distribution $\mathcal{D}_r$ that is inversely proportional to the distance. That is, let $u \in V$. For any $v \in V$, the probability to have a long link (u,v) is

$$\textit{Let } u \in V, \qquad P(u \to v) = \frac{1}{H_r(u)} \cdot \frac{1}{dist_G(u,v)^r} \qquad \textit{with } H_r(u) = \sum_{v \in V \setminus \{u\}} \frac{1}{dist_G(u,v)^r}$$

3.1 A bad solution: $r \neq D$. Example of the uniform distribution.

We consider the example of a 2-dimensional grid, where each node has one extra long link uniformly chosen among all vertices, i.e., $P(u \to v) = \frac{1}{n-1}$ $(r = 0)$.

Let $\epsilon < 1/4$. Let $t \in V$ be the destination of a message and let

$$B = \{u \in V \ : \ dist_G(u,t) \leq n^{\epsilon}\}.$$

Let p be the probability that there exists a vertex $u \in B$ with its long link going in B.

Lemma 1.

$$p = Prob\{\exists u, v \in B \ : \ (u,v) \in L\} \underset{n \to \infty}{\to} 0$$

Proof. Note that $|B| = \Theta(n^{2\epsilon})$. $p = 1 - Prob\{\forall u, v \in B \ : \ (u,v) \notin L\}$. Hence, $p = 1 - \Pi_{u \in B} Prob\{\forall v \in B \ : \ (u,v) \notin L\}$. So $p = O(1 - \Pi_{u \in B}(1 - \frac{|B|}{n})) = O(1 - (1 - \frac{1}{n^{1-2\epsilon}})^{n^{2\epsilon}})$.

$\ln[(1 - \frac{1}{n^{1-2\epsilon}})^{n^{2\epsilon}}] = n^{2\epsilon} \ln(1 - \frac{1}{n^{1-2\epsilon}}) = n^{2\epsilon}(-\frac{1}{n^{1-2\epsilon}} + o(\frac{1}{n^{1-2\epsilon}})) = -\frac{1}{n^{1-4\epsilon}} + o(\frac{1}{n^{1-4\epsilon}})$

Finally, $p = 1 - e^{-\frac{1}{n^{1-4\epsilon}} + o(\frac{1}{n^{1-4\epsilon}})} \underset{n \to \infty}{\to} 0$ (because $\epsilon < 1/4$). □

From previous lemma, when a message arrives at distance $\leq n^{\epsilon}$ to t, then no long links can be used to reach t. Hence,

Theorem 1. *[Kle00b] If $r = 0$ and $D = 2$, then the expected number of steps used by the greedy routing is at least $\Omega(n^{\epsilon})$, and $(G_2, \mathcal{D}_0)$ is not a small world.*

3.2 $r = D$ turns the grid into a small world.

We consider the augmented grid $(G_D, \mathcal{D}_r)$ with $r = D$, i.e., $P(u \to v) = \frac{1}{H_r(u)} \cdot \frac{1}{dist_G(u,v)^r}$. Let t be the destination node.

Theorem 2. *[Kle00b] In $(G_r, \mathcal{D}_r)$, the greedy routing performs in $O(\log^2 n)$ steps in expectation.*

Lemma 2. *Let $s \in V$ and $\delta = dist(s,t)$ and let $B = \{v \in V \ : \ dist(v,t) \leq \delta/2\}$.*

$$p = Prob\{\exists v \in B \ : \ (s,v) \in L\} = \Omega(\frac{1}{\log n}).$$

Proof. Note that $|B| = \Theta(\delta^r)$. Note also that the diameter of G is $rn^{1/r}$.

$H_r(s) = \sum_{v \in V \setminus \{s\}} \frac{1}{dist_G(s,v)^r} = \sum_{i=1}^{rn^{1/r}} |S_i|/i^r$ where $S_i = \{x \in V \ : \ dist(s,x) = i\}$. Since $|S_i| = \Theta(i^{r-1})$, we get that $H_r(s) = \Theta(\sum_{i=1}^{rn^{1/r}} 1/i) = \Theta(\log rn^{1/r}) = \Theta(\log n)$.

Let $v \in B$ such that $dist(v,s) = 3\delta/2$.

$p = \sum_{u \in B} Prob\{(s,u) \in L\} \geq |B| Prob\{(s,v) \in L\} = \frac{|B|}{H_r(s)} \cdot \frac{1}{(3\delta/2)^r} \geq \frac{\delta^r}{\log n} \cdot \frac{1}{(3\delta/2)^r} = \Theta(\frac{1}{\log n})$. □

Lemma 3. *Let $s \in V$ and $\delta = dist(s,t)$ and let $B = \{v \in V \ : \ dist(v,t) \leq \delta/2\}$. The expected number of steps to reach B from s is $O(\log n)$.*

Proof. Let $(s, x_1, x_2, \cdots)$ be the path followed by a message with destination t, according to the greedy routing. Since, for all $i \geq 1$, $dist(x_i, t) \leq dist(s,t)$, and because $\mathcal{D}_r$ is inversly proportional to the distance, we get that, for all $i \geq 1$, $Prob\{\exists v \in B \ : \ (x_i, v) \in L\} \geq \Omega(\frac{1}{\log n})$.

Bernoulli distribution: $Prob\{X = 0\} = p \leq 1$ and $Prob\{X = 1\} = 1 - p$. Then, the expected number of steps before getting a 0 is $\sum_{i \geq 1} i(1-p)^{i-1} p = 1/p$. □

Hence, in expectation, every $\log n$ steps, the greedy algorithm divides the distance from the current message's position to its destination by 2. So, it takes, in expectation, $\log n \log \ell$ steps for a message to reach its destination, where ℓ is the diameter of G_r, i.e., $\ell = rn^{1/r}$. This concludes the proof of Theorem 2.

Theorem 3. *[Kle00b] The D-dimensional grid with probability distribution $\mathcal{D}_r$ is a small world iff $D = r$.*

4 Beyond the grids: is every graph small-worldisable?

In previous sections, we saw how grids can be augmented into small worlds. That is, there is a probability distribution $\mathcal{D}$ such that the greedy routing algorithm performs in $\log^2 n$ steps (in expectation) in the D-dimensional grid augmented via $\mathcal{D}$. Apart the grid, several other graphs' classes have been investigated as bounded treewidth graph [Fra05], bounded growth graphs [DHLS06a], graphs excluding a minor [AG06], bounded doubling dimension metrics [Sli05], etc. In all these classes of graphs, probability distributions have been proposed to make the greedy routing algorithm to perform in poly-logarithmic number of steps. The question was to know whether similar probability distributions for any graph.

More generally, Question 1 can be reformulated as follows:

Question 2. What is the smallest function $f(n)$ such that there exists a probability distribution $\mathcal{D}$ such that the greedy routing algorithm performs in $f(n)$ steps (in expectation) in any graph augmented via $\mathcal{D}$?

Actually, it is easy to see that $f(n) = O(\sqrt{n})$ as noticed in [FGK$^+$09].

Lemma 4. $f(n) = O(\sqrt{n})$

Proof. Let G be any graph and, for any $v \in V$, choose uniformly at random $u \in V$ and add a long link (v, u). Now consider any target t. Consider the ball B of radius $\sqrt{n}$ centered at the target. For any vertex $x \notin B$, the probability that the long link (x, u) is such that $u \in B$ is $|B|/n = 1/\sqrt{n}$. Hence, the expected number of steps before the message arrives in B is $\sqrt{n}$. Once the message ais arrived in B, the number of remaining steps is at most $\sqrt{n}$. □

Moreover:

Theorem 4. *[FGK$^+$09]* $f(n) = O(n^{1/3})$.

Theorem 5. *[FLL10]* $f(n) = \Omega(2^{\sqrt{\log n}})$. *More precisely, there exists a infinite family of graphs such that for any augmentation scheme, greedy routing requires an expected number of* $\Omega(2^{\sqrt{\log n}})$ *steps, for some source-target pair.*

Theorem 6. *[FG10]* $f(n) = O(2^{\sqrt{\log n}}) = n^{o(1)}$.

5 Improving the greedy routing

Let $(G, \mathcal{D})$ be an augmented graph. In this section, we assume that each node knows the graph topology G, its own long links, but also some long links of some vertices of G. More precisely, it is natural to assume that a vertex knows the long links of the vertices that are close to it. We will see that this improves the performance of the greedy algorithm.

More precisely, let $v \in V$ and let K_v be the set of vertices u such that v knows the long links of w. Let $\bar{K}_v = \{w \in V \ : \ (u, w) \in V \ and \ u \in K_v\}$ be the set of head of the long links with tails in K_v.

The *indirect greedy routing* performed as follows. When a node v receives a message with destination $d \in V$, $d \neq v$, then v selects a vertex $u \in N_G(v) \cup N_{\mathcal{D}}(v) \cup \bar{K}_v$ such that the distance between u and d in G is minimum. Ties are broken uniformly at random. Then, v sends the message directly to u if $u \in N_G(v) \cup N_{\mathcal{D}}(v)$ or uses recursively the greedy routing to send the message to $w \in K_v$ such that $(w, u) \in L$ (in case $u \in \bar{K}_v$).

Remark. If at some step, a node v chooses to route the message toward a vertex u because v knows that u has a long link that will be closer to the destination, then during the route between v and u, then another node could know a better intermediary target and then the message will never go to u. Note that, depending on how the sets K_v are defined, the indirect greedy routing may have loops.

In what follows, K_v will be the set of the $\log n$ vertices closest to v. With such a setting, the indirect greedy routing converge (there are no loops).

Theorem 7. *[FGP06] For any* $v \in V$, *let* K_v *be the set of the* $\log n$ *vertices closest to* v. *Then, the undirect greedy routing performs in* $O(\log^{1+\frac{1}{r}} n)$ *steps in expectation in* $(G_r, \mathcal{D}_r)$.

Proof. Let $x \in V$ at distance δ to t. Let $B = \{v \in V \ : \ dist(v, t) \leq \delta/2\}$.

Let $u \in B$. $Prob\{\exists v \in K_x \ : \ (v, u) \in L\} = 1 - Prob\{\forall v \in K_x \ : \ (v, u) \notin L\} = 1 - \Pi_{v \in K_x} Prob\{(v, u) \notin L\} = 1 - \Pi_{v \in K_x}(1 - Prob\{(v, u) \in L\}) = 1 - \Pi_{v \in K_x}(1 - \frac{1}{\log n} \cdot \frac{1}{dist(v,u)^r})$.

If $\delta >> K_x^{1/r}$, then $\Pi_{v \in K_x}(1 - \frac{1}{\log n} \cdot \frac{1}{dist(v,u)^r}) = \Theta(1 - \frac{1}{\log n \cdot \delta^r})^{|K_x|}$.

So $Prob\{\forall v \in K_x : (v, u) \notin L\} = \Theta(e^{-\frac{|K_x|}{\log n \cdot \delta^r}})$, and

$$Prob\{\exists v \in K_x : (v, u) \in L\} = \Theta(1 - e^{-\frac{|K_x|}{\log n \cdot \delta^r}})$$

$Prob\{\exists v \in K_x, \exists u \in B \ : \ (v,u) \in L\} = \sum_{u \in B} Prob\{\exists v \in K_x \ : \ (v,u) \in L\}$. Hence,

$$Prob\{\exists v \in K_x, \exists u \in B \ : \ (v,u) \in L\} = |B|(1 - e^{-\frac{|K_x|}{\log n \cdot \delta^r}})$$

$Prob\{\exists v \in K_x, \exists u \in B \ : \ (v,u) \in L\} = \Theta(|B|\frac{|K_x|}{\log n \cdot \delta^r}) = \Theta(\frac{|K_x|}{\log n})$.

Therefore, in expectation, $\Theta(\frac{\log n}{|K_x|})$ sets K_x must be visited to find a long link toward B. Moreover, to cross each set K_x takes $|K_x|^{1/r}$ steps (the radius of K_x) in expectation. Hence, dividing the distance from the current message's position to its destination by 2 takes $\Theta(|K_x|^{1/r}\frac{\log n}{|K_x|})$ steps in expectation.

Therefore, to reach t takes $\Theta(\log n \cdot |K_x|^{1/r}\frac{\log n}{|K_x|})$ which is $O(\log^{1+\frac{1}{r}} n)$ for $|K_x| = \log n$. □

[LS05] proposed a decentralized routing algorithm that computes paths of length $O(\log n (\log\log n)^2)$, improving previous result. However, it is still larger than shortest paths ($O(\log n)$). Recently, Giakkoupis and Schabanel proved that:

Theorem 8. *[GS11] Consider the D-dimensional augmented grid.*

- *if $D \geq 2$, there is a decentralized routing algorithm that computes paths of expected length $O(\log n)$*
- *if $D = 1$, there is a decentralized routing algorithm that computes paths of expected length $O(\log n(\log\log n))$ and no distributed algorithm can do better.*

6 What's next? Applications...

The model of Kleinberg has been used for the design of several peer-to-peer protocols [GG02]. It is clearly interesting to understand how it could help for the design of routing tables in complex Networks like the AS-network of the Internet, etc.

References

[AG06] Ittai Abraham and Cyril Gavoille. Object location using path separators. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 188–197. ACM, 2006.

[CFL08] Augustin Chaintreau, Pierre Fraigniaud, and Emmanuelle Lebhar. Networks become navigable as nodes move and forget. In *Proceedings of the 35th International Colloquium on Automata, Languages and Programming (ICALP (1))*, volume 5125 of *Lecture Notes in Computer Science*, pages 133–144. Springer, 2008.

[DHLS06a] Philippe Duchon, Nicolas Hanusse, Emmanuelle Lebhar, and Nicolas Schabanel. Could any graph be turned into a small-world? *Theor. Comput. Sci.*, 355(1):96–103, 2006.

[DHLS06b] Philippe Duchon, Nicolas Hanusse, Emmanuelle Lebhar, and Nicolas Schabanel. Towards small world emergence. In *Proceedings of the 18th Annual ACM Symposium on Parallelism in Algorithms and Architectures (SPAA)*, pages 225–232. ACM, 2006.

[DMW03] Peter Dodds, Roby Muhamad, and Duncan Watts. An experimental study of search in global social networks. *Science*, 301(5634):827–829, 2003.

[FG09] Pierre Fraigniaud and George Giakkoupis. The effect of power-law degrees on the navigability of small worlds: [extended abstract]. In *Proceedings of the 28th Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 240–249. ACM, 2009.

[FG10] Pierre Fraigniaud and George Giakkoupis. On the searchability of small-world networks with arbitrary underlying structure. In *Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC)*, pages 389–398. ACM, 2010.

[FGK+09] Pierre Fraigniaud, Cyril Gavoille, Adrian Kosowski, Emmanuelle Lebhar, and Zvi Lotker. Universal augmentation schemes for network navigability. *Theor. Comput. Sci.*, 410(21-23):1970–1981, 2009.

[FGP06] Pierre Fraigniaud, Cyril Gavoille, and Christophe Paul. Eclecticism shrinks even small worlds. *Distributed Computing*, 18(4):279–291, 2006.

[FLL06] Pierre Fraigniaud, Emmanuelle Lebhar, and Zvi Lotker. A doubling dimension threshold *heta*(loglog) for augmented graph navigability. In *Proceedings of the 14th Annual European Symposium on Algorithms (ESA)*, volume 4168 of *Lecture Notes in Computer Science*, pages 376–386. Springer, 2006.

[FLL10] Pierre Fraigniaud, Emmanuelle Lebhar, and Zvi Lotker. Recovering the long-range links in augmented graphs. *Theor. Comput. Sci.*, 411(14-15):1613–1625, 2010.

[Fra05] Pierre Fraigniaud. Greedy routing in tree-decomposed graphs. In *Proceedings of the 13th Annual European Symposium on Algorithms (ESA)*, volume 3669 of *Lecture Notes in Computer Science*, pages 791–802. Springer, 2005.

[Fra07] Pierre Fraigniaud. Small worlds as navigable augmented networks: Model, analysis, and validation. In *Proceedings of the 15th Annual European Symposium on Algorithms (ESA)*, volume 4698 of *Lecture Notes in Computer Science*, pages 2–11. Springer, 2007.

[GG02] Hui Zhang 0002, Ashish Goel, and Ramesh Govindan. Using the small-world model to improve freenet performance. In *INFOCOM*, 2002.

[GS11] George Giakkoupis and Nicolas Schabanel. Optimal path search in small worlds: dimension matters. In *Proceedings of the 43rd ACM Symposium on Theory of Computing (STOC)*, pages 393–402. ACM, 2011.

[Kle00a] Jon M. Kleinberg. Navigation in a small world. In *Nature*, volume 406, page 845, 2000.

[Kle00b] Jon M. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proceedings of the 31st ACM Symposium on Theory of Computing (STOC)*, pages 163–170, 2000.

[Kle06] Jon M. Kleinberg. Complex networks and decentralized search algorithms. In *Proceedings of the International Congress of Mathematicians (ICM)*, 2006.

[LS05] Emmanuelle Lebhar and Nicolas Schabanel. Close to optimal decentralized routing in long-range contact networks. *Theor. Comput. Sci.*, 348(2-3):294–310, 2005.

[LS08] Emmanuelle Lebhar and Nicolas Schabanel. Graph augmentation via metric embedding. In *Proceedings of the 2th International Conference on Principles of Distributed Systems (OPODIS)*, volume 5401 of *Lecture Notes in Computer Science*, pages 217–225. Springer, 2008.

[Mil67] Stanley Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[MN04] Charles U. Martel and Van Nguyen. Analyzing kleinberg's (and other) small-world models. In *Proceedings of the Twenty-Third Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 179–188. ACM, 2004.

[NM05] Van Nguyen and Charles U. Martel. Analyzing and characterizing small-world graphs. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 311–320. SIAM, 2005.

[Sli05] Aleksandrs Slivkins. Distance estimation and object location via rings of neighbors. In *Proceedings of the Twenty-Fourth Annual ACM Symposium on Principles of Distributed Computing (PODC)*, pages 41–50. ACM, 2005.

[Wat99] D.J. Watts. Networks, dynamics, and the small-world phenomenon. *AJS*, 105(2):493–527, 1999.

[WS98] D.J. Watts and S.H Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:409–10, 1998.