8th Course: Diameter and center computations in networks

Cours MPRI 2010–2011

Michel Habib
habib@liafa.jussieu.fr
http://www.liafa.jussieu.fr/~habib

Chevaleret, november 2010
Schedule

Diameter Computations on Graphs and Networks

Structural explanations via graph theory
  K-chordal graphs
  Tree-length
  $\delta$-hyperbolic metrics
  Probabilistic analysis

Centers computations

Further work on diameter and center computations

Generalisation to suboptimal algorithms for graphs

Bibliography for diameter
Before I forgot

This is joint work with D. Corneil, V. Chepoi, F. Dragan, B. Estellon, M. Latapy, C. Magnien, C. Paul, Y. Vaxes
Basics Definitions

Definitions:
Let $G$ be an undirected graph:

- $\text{exc}(x) = \max_{y \in G} \{\text{distance}(x, y)\}$ excentricity
- $\text{diam}(G) = \max_{x \in G} \{\text{exc}(x)\}$ diameter
- $\text{radius}(G) = \min_{x \in G} \{\text{exc}(x)\}$
- $x \in V$ is a center of $G$, if $\text{exc}(x) = \text{radius}(G)$

First remarks of the definitions
distance computed in $\#$ edges
If $x$ and $y$ belong to different connected components $d(x, y) = \infty$.

diameter : Max Max Min
radius : Min Max Min
Trivial bounds

For any graph $G$:

\[ \text{radius}(G) \leq \text{diam}(G) \leq 2\text{radius}(G) \] and $\forall e \in G$,

\[ \text{diam}(G) \leq \text{diam}(G - e) \]

These bounds are tight

- If $G$ is a path of length $2K$, then $\text{diam}(G) = 2k = 2\text{radius}(G)$, and $G$ admits a unique center, i.e. the middle of the path.
- If $\text{radius}(G) = \text{diam}(G)$, then $\text{Center}(G) = V$. All vertices are centers (as for example in a cycle).
If $2 \cdot \text{radius}(G) = \text{diam}(G)$, then *roughly* $G$ has a tree shape (at least it works for trees).
But there is no nice characterization of this class of graphs.
Diameter

Applications

1. A graph parameter which measures the quality of services of a network, in terms of worst cases, when all have a unitary cost. Find critical edges $e$ s.t. $diam(G - e) > diam(G)$

2. Many distributed algorithms can be analyzed with this parameter (when a flooding technique is used to spread information over the network or to construct routing tables).

3. Verify the small world hypothesis in some large social networks, using J. Kleinberg’s definition of small world graphs.

4. Compute the diameter of the Internet graph, or some Web graphs, i.e. massive data.
FAQ

Usual questions on diameter, centers and radius:

► What is the best Program (resp. algorithm) available?
► What is the complexity of diameter, center and radius computations?
► How to compute or approximate the diameter of huge graphs?
► Find a center (or all centers) in a network, (in order to install serveurs).
Some notes

1. I was asked first this problem in 1980 by France Telecom for the phone network (FT granted a PhD).

2. Marc Lesk obtained his PhD in 1984 with the title: Couplages maximaux et diamètres de graphes. Maximum matchings and diameter computations

3. But, very little practical results.
Our aim is to design an algorithm or heuristic to compute the diameter of very large graphs.

Any algorithm that computes all distances between all pairs of vertices, complexity $O(n^3)$ or $O(nm)$. As for example with $|V|$ successive Breadth First Searches in $O(n(n + m))$.

Best known complexity for an exact algorithm is $O\left(\frac{n^3}{\log^2 n}\right)$.

At most $O(Diam(G))$ matrix multiplications.
1. Let us consider the procedure called: 2 consecutive BFS

**Data**: A graph $G = (V, E)$

**Result**: $u, v$ two vertices

Choose a vertex $w \in V$

$u \leftarrow \text{BFS}(w)$

$v \leftarrow \text{BFS}(u)$

*Where BFS stands for Breadth First Search.*

Therefore it is a linear procedure.
Intuition behind the procedure
Folklore

If $G$ is a tree, $diam(G) = d(u, v)$

Easy using Jordan’s theorem.
First theorem

Camille Jordan 1869:
A tree admits one or two centers depending on the parity of its diameter and furthermore all chains of maximum length starting at any vertex contain this (resp. these) centers.

And \( \text{radius}(G) = \lceil \frac{\text{diam}(G)}{2} \rceil \)
Unfortunately it is not an algorithm!
Certificates for the diameter

To give a certificate $\text{diam}(G) = k$, it is enough to provide:

- two vertices $x, y$ s.t. $d(x, y) = k$ ($\text{diam}(G) \geq k$).
- a subgraph $H \subset G$ with $\text{diam}(H) = k$ ($\text{diam}(G) \leq k$). $H$ may belong to a class of graphs on which diameter computations can be done in linear time.
Randomized BFS procedure

**Data**: A graph $G = (V, E)$

**Result**: $u, v$ two vertices

Repeat $\alpha$ times:

Randomly Choose a vertex $w \in V$

$u \leftarrow BFS(w)$

$v \leftarrow BFS(u)$

Select the vertices $u_0, v_0$ s.t. $distance(u_0, v_0)$ is maximal.
1. This procedure gives a vertex $u_0$ such that:
   \[ \text{exc}(u_0) \leq \text{diam}(G) \] i.e. a lower bound of the diameter.

2. Use a spanning tree as a partial subgraph to obtain an upper bound by computing its exact diameter in linear time.

3. Spanning trees given by the BFS.
The Program and some Data on Web graphs or P-2-P networks can be found

http://www-rp.lip6.fr/~magnien/Diameter

2 millions of vertices, diameter 32 within 1

Further experimentation by Crescenzi, Grossi, Marino ESA 2010

which confirm the excellence of the lower bound using LexBFS!!!!
Since $\alpha$ is a constant ($\leq 1000$), this method is still in linear time and works extremely well on huge graphs (Web graphs, Internet . . .)

How can we explain the success of such a method?

Due to the many counterexamples for the 2 consecutive BFS procedure. An explanation is necessary!
2 kind of explanations
The method is good or the data used was good.

Partial answer
The method also works on several models of random graphs.
So let us try to prove the first fact

Restriction
First we are going to focus our study on the 2 consecutive BFS.
Chordal graphs

1. A graph is chordal if it has no chordless cycle of length $\geq 4$.
2. If $G$ is a chordal graph, Corneil, Dragan, H., Paul 2001, using a variant called 2 consecutive LexBFS:
   $$d(u, v) \leq diam(G) \leq d(u, v) + 1$$
3. Generalized by Corneil, Dragan, Kohler 2003 using 2 consecutive BFS:
   $$d(u, v) \leq diam(G) \leq d(u, v) + 1$$
A nice algorithmic problem on subset families

Let $X$ be a finite set, and $\mathcal{F}$ be a family of subsets of $X$.

- Find a linear algorithm which computes if there exist $S, S' \in \mathcal{F}$ s.t. $S \cap S' = \emptyset$
- Linear i.e. in $O(|X| + |\mathcal{F}| + \sum_{S \in \mathcal{F}} |S|)$
Disjoint sets problem and diameter of split graphs
K-chordal graphs

[CDK’03]

If $G$ is $k$-chordal (i.e. $G$ does not contain any cycle of length $\geq k$), then the two consecutive BFS allow to find a vertex $x$ such that $ecc(x) \geq Diam(G) - \lfloor k/2 \rfloor$.

where $x$ is the middle of $[u, v]$. 
Diameter definition can be extended to any subset $A$ of vertices and let us denote by

$$\text{diam}(A) = \max_{x,y \in A} \{ d_G(x, y) \}$$

By convention $\text{diam}(\emptyset) = 0$.

Warning: distances are computed in the whole graph, not inside the subgraph $G(A)$!
Tree decomposition

Let us recall that a graph $G = (V, E)$ has a tree-decomposition $D = (S, T)$ if $S = \{S_1, S_2, \ldots, S_h\}$ is a collection of subsets of $V$, called bags, $T$ a tree whose vertices are elements of $S$ such that:

1. The union of elements in $S$ is $V$
2. $\forall e \in E$, $\exists i \in I$ with $e \in G(S_i)$.
3. $\forall x \in V$, the elements of $S$ containing $x$ form a subtree of $T$. 

Structural explanations via graph theory

Tree-length
An important property of tree decompositions:
Let $S_1 S_2$ be an edge of $T$ (joining the two bags $S_1$ and $S_2$), let $T_1$ and $T_2$ be the subtrees of $T$ obtained by removing the edge $S_1 S_2$. Then, $I = S_1 \cap S_2$ separates (i.e. is a separator in $G$) vertices of $T_1 - I$ from $T_2 - I$.

A very important notion in graph theory (B. Courcelle, P. Seymour, N. Roberston .....)

References
Tree-length by Dourisbourne and Gavoille 2003

Let us consider a new graph parameter, denoted by $treetlengh(G)$ and defined as follows:

$$Treetlengh(G) = \min_{\text{all } D} \{ \max_{S \text{ bag of } D} \{ \text{diam}(S) \} \}$$

In other words, for treewidth one measures the maximum size of a bag, as for tree-length one measures the maximum diameter of a bag.

- It is easy to see that: $Treetlengh(G) = 1$ iff $G$ is chordal.
- Obvious using the existence of a maximal clique tree
- If $G$ is a cograph then $Treetlengh(G) \leq 2$.
- An easy induction shows that any connected cograph has diameter $\leq 2$
- Idem for distance hereditary graphs.
Tree-length of known graphs

- Tree-length of $C_n$, the cycle of length $n$.
- It is easy to see that $TreeLength(C_n) \leq n/3$, by cutting the cycle into 3 bags of equal size and producing a triangulation of the cycle.
- Tree-length of the grid $G_{n,m}$
- Easy to show $TreeLength(G_{n,m}) \leq \min\{n, m\}$
Treewidth and Treelength are incomparable parameters.

- For a clique $K_n$, $\text{Treewidth}(K_n) = n-1 > \text{Treelength}(K_n) = 1$
- For a cycle of length $n$, $\text{Treewidth}(C_n) = 2 < \text{Treelength}(C_n) = n/3$
Tree-length $(G) \leq k$ iff there exist a chordal completion $H$ of $G$ in which all maximal cliques $C$ satisfies $diam_G(C) \leq k$.

Computing Treelength is NP-hard, D. Lokshtanov MFCS 2007!
Tree-length was introduced by Y. Dourisboure and C. Gavoille in order to capture the structure of a network. And they described some efficient routing protocols on networks having bounded treelength.

Chepoi, Dragan, H., Estellon, Vaxes 2007
If $G$ has Treelength $k$, then
\[2(\text{radius}(G) - k) \leq \text{diam}(G) \leq 2\text{radius}(G)\]
If $G$ has Treelength $k$, then $G$ is $k/2$-chordal.
Hyperbolic metric spaces

**Gromov’s 1987 Definition**

A graph is $\delta$-hyperbolic iff:
For every four vertices $u, v, w, z$ they are 3 distances (3 matchings) $d(u,v)+d(w,z)$ and $d(u,w)+d(v,z)$ and $d(u,z)+d(v,w)$

the two maximal values differ by at most $2\delta$

Gromov introduced this notion in the context of group theory.
A very interesting definition

Which is both local and global.
Local for the 4 points, global by the distance in the whole graph.

$\delta$-hyperbolicity can be easily computed in $O(n^4)$, therefore it is polynomial to compute.
Misha Gromov from his web page! Abel Prize 2009
Structural explanations via graph theory
\( \delta \)-hyperbolic metrics

Misha Gromov
For a tree $\delta = 0$
For general graphs

general case
Trivial examples

- For trees:
  If some edge $e$ split the vertices in two parts $:u, v$ and $x, y$, then the two sums that contain $e$ are the biggest ones and equal.
  Else the three sums are equal.
- For a cycle of length $2k$, $\delta$-hyperbolicity is $\frac{k}{2}$
2 geodesics
2 geodesics

**Lemma**

Let us consider 2 geodesics from $u$ to $v$. If $x$ and $y$ travel at the same speed from $u$ to $v$ on these geodesics

Then $d(x, y) \leq 2\delta$

**Proof**

Considering the three sums on these four points we have :

$$d(x, u) + d(y, v) = d(x, v) + d(y, u) = d(u, v) \leq d(x, y) + d(u, v)$$
Why this notion is so interesting?

1. A metric space embeds into a tree iff for any four points the two larger sums are equal. ($\delta = 0$).
2. $\delta$-hyperbolicity is a kind of measure via metric distances to a tree.
3. Many usual graph classes have small $\delta$-hyperbolicity.
- $\delta(K_n) = 0$, $\delta(G) = 0$ iff $G$ is a cactus of cliques.
- $G$ chordal implies $\delta(G) = 1$.
- Chepoi characterized graphs such that $\delta(G) = 1$.
- $\delta(G) \leq 2$ if $G$ is AT-free.
If $G$ has treelength $k$ then $G$ is $k$-hyperbolic

For a $\delta$-hyperbolic graph, $d(u, v) \geq \text{diam}(G) - 2\delta$
and $\text{diam}(C(G)) \leq 4\delta + 1$
We obtain (2008) for $\delta$-hyperbolic graphs, generalizing Gavoille’s results on treelength:

- An additive $O(\delta \log n)$-spanners with at most $O(\delta n)$ edges
- An $O(\delta \log n)$-additive routing labeling scheme with $O(\delta \log^2 n)$ bit labels.
Graphs can be transformed in a metric space replacing edge by a segment of length 1. Same theorems for discrete metrics and usual ones.

So we can apply our results to geometric graphs such as polygons and the results are still valid. Let P be a simple polygon, LG(P) its link graph, we prove it has \( \delta \)-hyperbolicity 2.

Relationships with our 2 BFS method and some method in computational geometry to obtain the center of a polygon.
Some open questions for $\delta$-hyperbolicity

- Provide efficient computations or approximations for the $\delta$-hyperbolicity.
- $\delta(G) = 0$ can be recognized in $O(n + m)$
- $\delta(G) = 1$ Using Chepoi characterization?
- $\delta(G) = 2$?
- $\delta(G) = 3$?
Further questions

1. Changing the definition: multiplicative or max in the four point condition.
2. What remain correct on the weighted case?
3. Does there exists a general meta-theorem for graphs of bounded $\delta$-hyperbolicity?
4. What are the critical points, relatively to the 4 points condition?
Other applications of Gromov’s framework

Phylogenetics in Biology.
Gromov’s operation which transforms an unrooted hierarchy into a rooted one.
Lior Patcher’s recent work
Back to the diameter problem what have we obtained so far?

- If the graph is closed to a tree (bounded cycles for K-chordality)
- Has a tree structure (treelength parameter)
- or closed to a tree in a metric way for small $\delta$-hyperbolicity
- then we can bound the behavior with an additive constant of the 2-consecutive BFS.
- We should go further ....
First Attempt

- Parnas and Ron 2004, they study the efficiency of a single BFS to test the diameter.
- The starting vertex is chosen at random.
- So the probabilistic analysis on the 2-sweep heuristic remains to be done!
Centers for chordal graphs

- Let $C$ be the set of all centers of $G$, if $G$ is chordal then $G(C)$ is m-connected, $diam(G(C)) \leq 3$ and $radius(G(C)) \leq 2$.
- Chepoi and Dragan ESA 1994 used this property to build a beautiful linear time algorithm to find a center in a chordal graph.
- They generalized this technique to other classes of graphs: HDD-free . . .
  For k-chordal, $diam(G(C)) \leq k$. 
Center computations

- Is there a computational difference between diameter and center?
- Computation of centers seems to be easier (at least for chordal graphs).
- Can you compute $C(G)$ the whole set of centers in linear time for chordal graphs?
Real Data ? from CAIDA project

M. Soto, PhD student at Paris Diderot, has computed graph invariants on some real networks

2 graphs with normal graph distance

Internet Topology Data Kit (ITDK) graph of the routing machines
Treedwidth $\geq 234$, Treelength $\leq 10$, Diameter $= 19$, $\delta$-hyperbolicity $= 3$ (but for 96 % of the computed quadruplets the value is 1)

Autonomus System Internet Topology (AS-level) graph, a smaller graph
Treedwidth $\geq 82$, Treelength $\leq 6$, Diameter $= 10$, $\delta$-hyperbolicity $= 2$ (but for 98 % of the computed quadruplets the value is 1)
First remarks

- usual metric for graphs, distance = length of the path = \# number of edges
- It is far beyond the scope of our actual knowledge to compute the treewidth of such graphs (20 000 vertices)
- $\delta$-hyperbolicity seems to be an interesting parameter for networks. (already noticed by R. Kleinberg and others).
Further work on diameter and center computations

Research directions

- Generalize the BFS approach to computations of radius and centers.
- Extend the probabilistic analysis of the 2 consecutive BFS procedure.
- We still do not know precisely if the computation of the diameter requires the computation of all distances.
- Non uniform diameter (weighted graphs)
- Dynamic maintenance of diameter
- Distributed versions for networks.
Suboptimal algorithms

- The idea is to only parse part of the input to obtain evaluations of the property you want to compute. Also known as property testing.
- When can you say that a graph is not bipartite with high probability, without considering the whole graph. Many Noga Alon’s papers on these kind of questions.
The maximal flow problem: a good case study

- Many polynomial algorithms available, but not easy to use on massive graphs.
- No linear-time approximation known.
- Strangely, such approximation algorithms are known for many NP-complete problems!
What is known

- Max flow $=$ Min cut theorem.
- It seems to be easier to find a good cut than a good flow.
Hints to attack the problem

- PageRank obtained by matrix computations on huge Web Graphs
- Use PageRank to build a kind of preflow . . .
- Need for 20 years of hard research?
- Or just 20 years to find the right idea?
This approach can be used for all polynomial problems for graphs for which the exact algorithms have a complexity quadratic or $O(n.m)$. Somehow the complexity barrier of the boolean matrix multiplication.
Succint Bibliography for diameter

- T. M. Chan, All-pairs shortest paths for unweighted undirected graphs in $o(mn)$ time, SODA 2006.
- D. Kratsch, J. Spinrad, Between $O(mn)$ and $O(n^\alpha)$, SODA 2003.
Bibliography continued

- V. Chepoi, F. Dragan, B. Estellon, M. Habib, Y. Vaxès, Diameters, centers, and approximation trees of $\delta$-hyperbolic spaces and graphs, ACM Computational Geometry Conf. 2008
Thank you for your attention during these 8 courses!!

Merci de votre attention durant cette première partie du cours!!