A Note on the Complexity of Computing the Number of Reachable Vertices in a Digraph

Michele Borassi

March 26, 2018

Abstract

In this work, we consider the following problem: given a digraph $G = (V, E)$, for each vertex v, we want to compute the number of vertices reachable from v. In other words, we want to compute the out-degree of each vertex in the transitive closure of G. We show that this problem is not solvable in time $\mathcal{O}\left(|E|^{2-\epsilon}\right)$ for any $\epsilon > 0$, unless the Strong Exponential Time Hypothesis is false. This result still holds if G is assumed to be acyclic.

1 Introduction

In this work, we consider the following problem: given a digraph $G = (V, E)$, for each vertex v, we want to compute the number of vertices reachable from v. An efficient solution of this problem could have many applications: to name a few, there are algorithms that need to compute (or estimate) these values [6], the number of reachable vertices is used in the definition of other measures, like closeness centrality [10, 14, 11], and it can be useful in the analysis of the transitive closure of a graph (indeed, the out-degree of a vertex v in the transitive closure is the number of vertices reachable from v).

Until now, the best algorithms to solve this problem explicitly compute the transitive closure of the input graph, and then output the out-degree of each node. This can be done through fast matrix multiplication [8], in time $\mathcal{O}(N^{2.373})$ [16], or by performing a Breadth-First Search from each node, in time $\mathcal{O}(MN)$, where $N = |V|$ and $M = |E|$.

However, one might think that if only the number of reachable vertices is needed, then there might be a faster algorithm: in this work, we prove that this is not the case, even if the input graph is acyclic. Indeed, an algorithm running in time $\mathcal{O}(M^{2-\epsilon})$ would falsify the well-known Strong Exponential Time Hypothesis [9]: this hypothesis says that, for each $\delta > 0$, if k is big enough, the k-Satisfiability problem on n variables cannot be solved in time $\mathcal{O}((2-\delta)^n)$. As far as we know, this reduction has never been published, even if several similar reductions are available in the literature [15, 17, 12, 13, 4, 2, 7, 1, 5, 3].

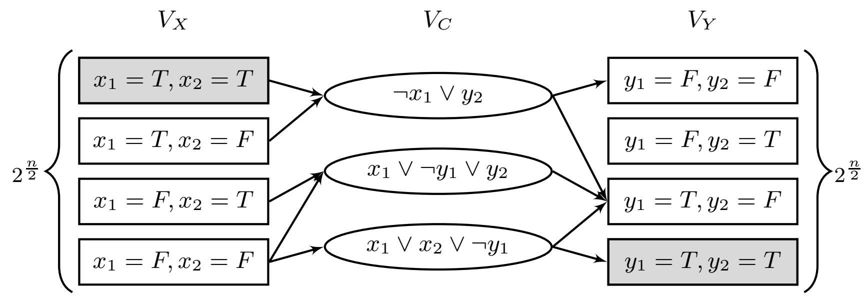

Figure 1: An example of the graph obtained from the formula $(\neg x_1 \vee y_2) \wedge (x_1 \vee \neg y_1 \vee y_2) \wedge (x_1 \vee x_2 \vee \neg y_1)$. The two gray evaluations correspond to a satisfying assignment.

2 The Reduction

Let us consider an instance of the k-Satisfiability problem on n variables, and let us assume that $n = 2l$ (if n is odd, we add one variable that does not appear in any clause). Let us divide the variables in two sets X, Y, such that $|X| = |Y| = l$. We will name $x_1, \dots, x_l$ the variables in X, and $y_1, \dots, y_l$ the variables in Y. From this instance of the k-Satisfiability problem, let us construct a digraph as follows. We consider the set V_X of all $2^{|X|}$ possible evaluations of the variables in X, and the set V_Y of all $2^{|Y|}$ possible evaluations of the variables in Y. The set of vertices is $V_X \cup V_Y \cup V_C$, where V_C is the set of clauses.

We add a directed edge from a vertex $v \in V_X$ to a vertex $w \in V_C$ if the evaluation v does not make the clause w true. For instance, if $X = \{x_1, x_2\}$, and w is $x_1 \vee y_2$, the evaluation $(x_1 = T, x_2 = T)$ is not connected to w, because the variable x_1 makes the clause w true. Conversely, the evaluation $(x_1 = F, x_2 = T)$ is connected to w, because it does not make the clause w true (note that, in this case, we still can make w true by setting $y_2 = T$). Similarly, we add a directed edge from a clause $w \in V_C$ to an evaluation v in V_Y if the evaluation v does not make the clause w true. An example is shown in Figure 1.

The formula is satisfiable if and only if we can find an evaluation $v_X \in V_X$ of the variables in X and an evaluation $v_Y \in V_Y$ of the variables in Y such that each clause is either satisfied by v_X or by v_Y. By construction, this happens if and only if v_X is not connected to v_Y in the graph constructed (for example, the two gray evaluations in Figure 1 correspond to a satisfying assignment).

Moreover, the graph constructed has at most $N = |X| + |Y| + |C| \leq 2 * 2^{\frac{n}{2}} + n^k = \mathcal{O}\left(2^{\frac{n}{2}}\right)$ nodes, and at most $M = |X||C| + |Y||C| \leq 2 * 2^{\frac{n}{2}} * n^k = \mathcal{O}\left(2^{\frac{n}{2}} n^k\right)$ edges.

This means that, if we can count the number of reachable vertices in time $\mathcal{O}(N^{2-\epsilon})$, then we can also verify if the formula is satisfiable, by checking if all vertices in V_X can reach all vertices in V_Y with no overhead (the number of vertices in V_Y reachable from a vertex $v \in V_X$ can be computed in time

$\mathcal{O}(n^k)$ as the total number of vertices reachable from v, minus the number of vertices in V_C reachable from v). As a consequence, if we have an algorithm that computes the number of reachable vertices in time $\mathcal{O}(M^{2-\epsilon})$ for some ϵ, then we can find an algorithm that solves k-SATISFIABILITY in time $\mathcal{O}\left(\left(2^{\frac{n}{2}}n^k\right)^{2-\epsilon}\right) = \mathcal{O}\left(\left(2^{\frac{2-\epsilon}{2}}\right)^n n^{(2-\epsilon)k}\right) = \mathcal{O}\left((2-\delta)^n\right)$ for a suitable choice of δ. This falsifies the Strong Exponential Time Hypothesis, and concludes the reduction.

Acknowledgements

The author thanks Emanuele Natale for reading carefully and correcting the first version. He also thanks Emanuele Natale and Massimo Cairo for suggesting him to write the short paper.

References

[1] Amir Abboud, Fabrizio Grandoni, and Virginia V. Williams. Subcubic equivalences between graph centrality problems, APSP and diameter. In *Proceedings of the 26th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2015.

[2] Amir Abboud and Virginia V. Williams. Popular conjectures imply strong lower bounds for dynamic problems. In *Proceedings of the 55th Annual Symposium on Foundations of Computer Science (FOCS)*, 2014.

[3] Amir Abboud, Virginia V. Williams, and Joshua Wang. Approximation and Fixed Parameter Subquadratic Algorithms for Radius and Diameter. In *Proceedings of the 27th ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2016.

[4] Amir Abboud, Virginia V. Williams, and Oren Weimann. Consequences of Faster Alignment of Sequences. In *Proceedings of the 41st International Colloquium on Automata, Languages and Programming (ICALP)*, 2014.

[5] Michele Borassi, Pierluigi Crescenzi, and Michel Habib. Into the square - On the complexity of some quadratic-time solvable problems. In *Proceedings of the 16th Italian Conference on Theoretical Computer Science (ICTCS)*, 2015.

[6] Michele Borassi, Pierluigi Crescenzi, Michel Habib, Walter A. Kosters, and Frank W. Takes. Fast Diameter and Radius BFS-based Computation in (Weakly Connected) Real-World Graphs - With an Application to the Six Degrees of Separation Games. *Theoretical Computer Science*, 2014.

[7] Karl Bringmann. Why walking the dog takes time: Frechet distance has no strongly subquadratic algorithms unless SETH fails. In *Proceedings of*

the 55th Annual Symposium on Foundations of Computer Science (FOCS), 2014.

[8] Michael J. Fischer and Albert R. Meyer. Boolean matrix multiplication and transitive closure. In *Proceedings of the 12th Annual Symposium on Switching and Automata Theory*, 1971.

[9] Russell Impagliazzo, Ramamohan Paturi, and Francis Zane. Which Problems Have Strongly Exponential Complexity? *Journal of Computer and System Sciences*, 2001.

[10] Nan Lin. *Foundations of social research.* McGraw-Hill, 1976.

[11] Paul W. Olsen, Alan G. Labouseur, and Jeong-Hyon Hwang. Efficient top-k closeness centrality search. *Proceedings of the IEEE 30th International Conference on Data Engineering (ICDE)*, 2014.

[12] Mihai Pătraşcu and Ryan Williams. On the possibility of faster SAT algorithms. In *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2010.

[13] Liam Roditty and Virginia V. Williams. Fast approximation algorithms for the diameter and radius of sparse graphs. *Proceedings of the 45th ACM Symposium on Theory of Computing (STOC)*, 2013.

[14] Stanley Wasserman and Katherine Faust. *Social Network Analysis: Methods and Applications.* Structural Analysis in the Social Sciences. Cambridge University Press, 1994.

[15] Ryan Williams. A new algorithm for optimal 2-constraint satisfaction and its implications. *Theoretical Computer Science*, 348, 2005.

[16] Virginia V. Williams. Multiplying matrices faster than Coppersmith-Winograd. In *Proceedings of the 44th ACM Symposium on Theory of Computing (STOC)*, 2012.

[17] Virginia V. Williams and Ryan Williams. Subcubic Equivalences between Path, Matrix and Triangle Problems. In *Proceedings of the 51st Annual Symposium on Foundations of Computer Science (FOCS)*, October 2010.