

Electronic Design Automation for Social Networks

Andrew DeOrio and Valeria Bertacco

Electrical Engineering and Computer Science, University of Michigan
{awdeorio, valeria}@umich.edu

ABSTRACT

Online social networks are a growing internet phenomenon: they connect millions of individuals through sharing of common interests, political and religious views, careers, *etc.* Social networking websites are observing an ever-increasing number of regular users, who rely on this virtual medium to connect with friends and share in the community. As a result, they have become the repository of a vast amount of demographic information, which could deliver valuable insights to businesses and individuals. However, as of today, this data is for the most part still untapped, partly because of the complexity entailed by analyzing some of these vast social connectivity graphs. Another area that deals with large data sets is Electronic Design Automation (EDA), the result of increasingly complex computer systems. The powerful tools used to deal with these data sets open many possibilities for social networks. In this work we propose to study interesting aspects of social networks by deploying some of the solutions commonly used in EDA.

Categories and Subject Descriptors

B.6 [Logic Design]: Design Aids; J.4 [Computer Applications]: Social and Behavioral Sciences

General Terms

Human Factors, Algorithms, Verification

Keywords

Social Networks, EDA Algorithms, Verification

1. INTRODUCTION

The popularity of online social networking has exploded in recent years, reaching over 400 millions members in some instances (*e.g.* Facebook [6]). Monthly visits routinely exceed 1 billion for Facebook, with Twitter and MySpace in the range of 10 and 100 million [1]. With so many users and vast numbers of interactions, the information stored in these sites has significant potential, in business (for instance, in identifying ideal target groups for business growth), as a source for market data analysis (such as voting polls or product market-share) and in public safety investigations. However, as of today, this potential remains mostly unrealized, partly because it is not clear yet how it can be leveraged effectively [7], and because the scope of social networks is still limited, for the most part, to providers of social entertainment and communication channels between individuals.

Another industry that deals with large data sets is Electronic Design Automation. The design of a modern computer chip is a complex process and many specialized software tools have been developed to automate the process, enabling engineers to tackle ever-increasing complexity. Pairing computer aided design (CAD) techniques of the EDA industry with data analysis problems facing the social networking industry opens a myriad of possibilities.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

DAC'10, July 13 - 18, 2010, Anaheim, California, USA.

Copyright 2010 ACM 978-1-4503-0002-5 /10/06 ...\$10.00.

For example, the problem of verifying the validity of information about a user across different social networking sites rings of subgraph isomorphism, an EDA technique used to determine if a design produces the same functionality after an optimization. Subgraph isomorphism might also be useful for matching users with each other and with opportunities such as match making, job hunting and community organizing. Another fascinating possibility of applying EDA to social networks is in identifying and analyzing communities of users rather than individual users' own traits. This might be beneficial for advertising similar products to a group, with the group boundary identified by tools such as min-cut, which partitions a graph based on the number of connections between the resulting subgraphs.

Another possible application of EDA tools to social networking is model checking, which uses specialized formalisms and logics to reason about graphs, including CTL, LTL [4], together with clever solutions to cope with their size, such as abstraction [5] and efficient computation engines [8]. In the following section, we show that the application of model checking in the context of social networks has the potential to extract valuable insights. The ability to leverage these repositories, in some or all of the above domains, can bring social networks beyond the stage of a sensation to a central component in the fabric and operation of modern society.

2. CASE STUDY: MODEL CHECKING

One of the mainstream tools used by the EDA community to analyze complex graphs is model checking, a formal verification technique used to rigorously prove properties of a hardware design. Model checking is apt for analyzing social networks because of its ability to reason about a data set using properties. Properties are formal descriptions of a particular functionality of a design and are verified by the model checker against an internal mathematical representation of the input design.

We now show how traditional model checking solutions can be applied to social networks. As a first step, we note that classic social network studies represent people as nodes and their relationships as edges in a graph. Figure 1 shows an example of such graphs. The information reported is similar to what can be found on sites such as Twitter [3]: each vertex is a person or a virtual entity capable of communicating (*e.g.* , a business) participating in the social network. An edge connects user A to user B if A *follows* B, that is, if A can receive messages from B. For instance, with reference to the picture, if Maria broadcasts a message, or *tweets*, then Dick and Pierre will receive that message. Note that information flows in the opposite direction of the edges' vectors.

Nodes frequently have additional information associated with them. In the example we annotated each vertex with the country affiliation of the entity represented by each node.

Using Computational Tree Logic (CTL), we can reason about this graph by expressing and verifying properties. For example, if the White House were to send an important Twitter message, will everyone in the network potentially receive the message (assuming intermediate nodes decide that the message is sufficiently important to be forwarded (re-tweeted)? This can be expressed in CTL as $x \models \mathbf{EF} \text{ White_House}$, where x is the set comprising all the nodes in the graph. In the small context of our example this property is

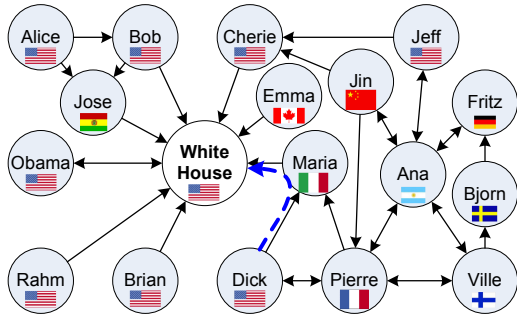


Figure 1: Example of a social network graph, a small subset of the type of information that can be gathered from the Twitter social network. Vertices represent users and edges highlight who they are following, with information flowing in the opposite direction of the arrows. The dashed bold arrow highlights a counterexample for the false property *messages from the White House may reach all US-based users exclusively through US-based users' re-tweets*.

indeed valid, since there is a path from each vertex to the White House. We could also restrict the property by simply asking if *all US-based users may receive the White House's messages (tweets)*:

$$x \models \text{location}=\text{US} \rightarrow \mathbf{EF} \text{ White_House}$$

which is also a true property, easily derived from the validity of our previous property. Finally, we extend our example by checking if all US-based users in the network may receive a tweet from the White House exclusively via other US-based re-tweets. In CTL, this final property can be expressed as:

$$x \models \text{location}=\text{US} \ \& \ \mathbf{E}(\text{location}=\text{US} \ \mathbf{U} \ \text{White_House})$$

This latter property is false: for instance, Dick cannot receive tweets from the White House unless Maria or Pierre broadcast them. A dashed bold arrow highlights the counterexample in Figure 1.

Analyzing Twitter. As an exploratory exercise, we collaborated with a local Twitter application startup [2], applying formal verification techniques to reason about its customer base. First, it was necessary to generate a graph representing a subset of the Twitter's user population which included the customers of the startup and its social surrounding (users connected to customers in either tweeting direction). We accomplished this with a PHP script interfacing with the Twitter API that gathered user information and relationships. The resulting graph had 46,782 vertices (users) and 68,562 edges. In addition to follower relationships, we also annotated the graph with the time and date of when each customer signed up for the service, assigning 0 to non-customers. This attribute is denoted with *date* in our discussion below.

We then investigated a number of properties in this graph by applying the model checking tool NuSMV [4]. The goal of the first property we considered was to determine if there were any customers connected solely to other customers, expressed in CTL as:

$$P0: x \models (\text{date} > 0) \rightarrow \mathbf{AX} (\text{date} > 0)$$

where *x* includes all the users in our sub-network. The model checker proved this to be incorrect with a counterexample. Similar properties included:

$$P1: x \models (\text{date} > 0) \rightarrow \mathbf{AX} (\text{date} \leq 0)$$

$$P2: x \models (\text{date} \leq 0) \rightarrow \mathbf{AX} (\text{date} > 0)$$

$$P3: x \models (\text{date} \leq 0) \rightarrow \mathbf{AX} (\text{date} \leq 0)$$

We then attempted to study more closely specific customers. Is there a tweeting path from Alice to Bob?

$$P4: x \models \text{user}=\text{Alice} \rightarrow \mathbf{EF} \text{ user}=\text{Bob}$$

Can Alice reach a non-customer?

$$P5: x \models \text{user}=\text{Alice} \rightarrow \mathbf{EF} (\text{date} \leq 0)$$

Can Alice's tweets return to herself exclusively via other customers?

$$P6: x \models \text{user}=\text{Alice} \rightarrow \mathbf{E}[(\text{date} > 0) \ \mathbf{U} \ \text{user}=\text{Alice}]$$

We verified all the above properties with NuSMV and reported the results in Table 1. As it can be gathered, all problems were easily handled by the model checker, which was able to provide results within approximately 18 minutes and using 160MB of memory.

Property	BDD nodes	time (s)	memory (MB)	result
P0	1,225,484	1,124	160	false
P1	1,362,742	1,124	161	false
P2	1,329,483	1,114	161	false
P3	1,269,974	1,127	161	false
P4	1,297,351	1,007	144	true
P5	1,346,374	1,136	155	true
P6	1,184,973	1,130	155	true

Table 1: Social properties model checked on the Twitter sub-network comprised of the startup's customers.

Scaling. While our study focused a subset of the Twitter network with less than 50,000 vertices, spanning the entire network would be much more challenging. The full Twitter base is estimated in the 10's of millions connected by billions of edges. We expect that the full scale problem would require aggressive abstraction and abstraction refinement techniques of such as those proposed in the past decade by the formal verification community. Given the typical complexity that these advanced methods can often tackle, we believe that they have the potential to let us investigate complex properties on some of the largest social networks.

3. CONCLUSIONS AND FUTURE WORK

The popularity of online social networking presents a myriad of opportunities for businesses, governments and individuals. The great quantities of information encapsulated in these networks could benefit from powerful analysis techniques common in the EDA industry. We proposed applying powerful tools used to handle large data sets in the EDA industry to social networks.

With a model checking example on a subset of the Twitter graph, we outlined the type of investigations that would be possible in this context. Moving forward, we hope to expand our studies to larger social networks, such as the entire Twitter system, and Facebook. Additionally, we would like to investigate the benefits that other EDA tools can bring to complex social network analysis.

Acknowledgments: We would like to thank Kyle Mulka of Twilk Inc. [2] for his assistance with the Twitter data collection.

4. REFERENCES

- [1] Social networks: Facebook takes over top spot, twitter climbs. *Compete*, 9 January 2009. blog.compete.com.
- [2] Twilk incorporated, 2010. twilk.com.
- [3] Twitter incorporated, 2010. twitter.com.
- [4] A. Cimatti, E. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. NuSMV Version 2: An OpenSource Tool for Symbolic Model Checking. In *Proc. CAV*, 2002.
- [5] E. Clarke, O. Grumberg, S. Jha, Y. Lu, and H. Veith. Counterexample-guided abstraction refinement for symbolic model checking. *50(5):752-794*, 2003.
- [6] Facebook. Facebook statistics, 2010. facebook.com/press/info.php?statistics.
- [7] Russell Herder and Ethos Business Law. Social media: Embracing the opportunities, averting the risks. In *Russell Herder and Ethos Business Law*, 2009.
- [8] L. Zhang and S. Malik. *The Quest for Efficient Boolean Satisfiability Solvers*. Springer Berlin / Heidelberg, 2002.