Team success in the iGEM scientific competition

Marc Santolini[1,2], Abhijeet Krishna, Christos Ellinas, Leo Blondel, Thomas E. Landrain, and Albert-Laszló Barabásí[1,2]

[1] Northeastern University, Boston MA
[2] Harvard Medical School, Boston MA
m.santolini@neu.edu

Abstract. This work investigates criteria of performance and success of teams in a scientific context. We leverage laboratory notebooks edited on wiki websites by student teams participating to the international Genetically Engineered Machines (iGEM) synthetic biology competition to uncover what features of team work best predict short term quality (medals, prizes) and long term impact (how the biological parts that teams engineer are re-used by other teams). This represents a large-scale dataset of 2,000 teams over 10 years, with an average 10 students per team, providing an unprecedented insight into the making of science.

1 Introduction

Recently, the literature has bloomed in large scale analyses of science as an object of study, paving the way for the "Science of science". Pervasive to this literature is the use of networks. Indeed, research is a collective phenomenon where any finding relies on previous knowledge elaborated by others [1]. The co-authorship produces a collaboration network informative on the social biases of research [2, 3], and the citation network allows to measure the impact and spread of new concepts [4].

While there is a trove of large scale datasets relative to the outputs of science, much less can be said with regards to the making of science *in situ*, in the laboratory. At the qualitative level, social scientists have been investigating this question decades ago, with early work by Latour and Woolgar [5] exhibiting the anthropological aspects of making science. Yet, such investigations have been lacking a quantitative counterpart, in part due to technical limitations. The necessary toolkit is nonetheless ready. For example, team metrics and their relation to team success have been measured in collaborative coding [6], in the artistic setup of Broadway musical [3], or in private organisations [7, 8]. Here we explore their role in the context of scientific production.

We leverage the iGEM Competition[3] of Synthetic Biology. For over 10 years, iGEM has been encouraging students to work together to solve real-world challenges by building genetically engineered biological systems with standard, interchangeable parts or BioBricks. Student teams design, build and test their projects over the summer and gather to present their work and compete at the annual Jamboree. A condition of participation to iGEM is that teams document their progress

[3] http://igem.org/Main_Page

and results on an open wiki website[4]. Given the underlying structure of wikis, it is possible to know which team member has edited which part of the wiki at what time. Finally, teams are awarded medals and special prizes (short term impact), and the BioBricks that they engineer can be later re-used by other teams in later years (long term impact). In this work, we investigate how features of team organization (obtained through the wiki) affect team success (medals, prizes etc) in this scientific context.

2 Results

We extracted team information at multiple levels, as shown in Figure 1. First, we built a scraper to extract the wiki history and content for 1,551 teams, information that was used to build internal team interaction networks. For each team, a bipartite network was constructed between the wiki editors (the team students) and the sections edited. Team networks were then reconstructed by projecting the bipartite network on the user space, counting the number of wiki subsections co-edited by any two students of a team. The obtained number was compared to the expected co-edition resulting from a hypergeometric distribution and a Z-score was computed. Finally, edges with $Z > 2$ were deemed significant and kept for further analyses. Teams also collaborate with one another, and we extracted for each year the team collaboration network. Teams produce BioBricks, and we extracted the number of BioBricks produced and their re-use. Finally, success measures were collected, consisting of the type of medal (None, Bronze, Silver or Gold), number of special prizes, being a Finalist and being a Winner of the competition.

Analysis of the data showed higher productivity per capita (number of edits, number of sections edited, number of bioparts produced) and larger time invested in the project for teams winning higher quality medals. Moreover, we observed that teams with higher network density, largest connected component, and a leader with high degree were significantly more successful in the competition. Finally, we built a classifier combining these different features together by fitting a logistic regression model with cross-validation. The classifier was able to predict winning teams with an Area under the ROC Curve $AUC = 0.8$, showing the high degree of relevance of the captured features with the observed outcome.

3 Discussion and future prospects

The iGEM competition offers a model system to understand the making of science in a controlled context. Here we have leveraged the publicly available data from the competition by making use of the open wiki websites. Future prospects consist in exploring the finer-grained judging data currently being obtained from the iGEM HQ. For each team, 6 judges rate 60 criteria from 1 to 5, quantifying project creativity and quality[5]. By accessing this data, we will have an unprecedented insight into how science quality is assessed, the variability of the assessment between judges, as well as what features correlate with more subjective metrics such as project creativity.

[4] To see an example wiki: http://2016.igem.org/Team:LMU-TUM_Munich
[5] http://2017.igem.org/Judging/Rubric

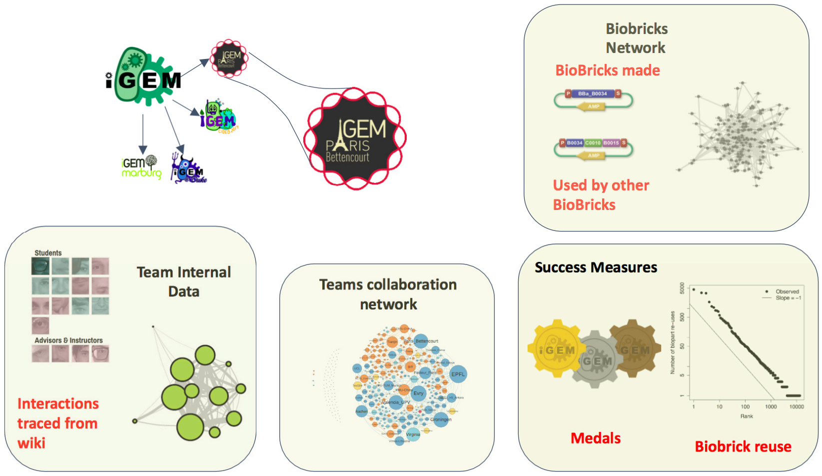

Fig. 1. Overview of the dataset. Over 10 years, 2,000+ teams have participated to the iGEM competition. Team internal networks are reconstructed from wiki notebooks co-edition. Teams collaborate with one another, forming a collaboration network. They produce BioBricks by combining previously made BioBricks or engineering new ones. Finally, team success is determined by the prizes and medals they receive, as well as their BioBricks re-use by other teams.

References

1. Sinatra,R., Deville,P., Szell,M., Wang,D. and Barabasi,A.-L. (2015) A century of physics. Nature Physics, 11, 791–796.
2. Shen,H.-W. and Barabasi,A.-L. (2014) Collective credit allocation in science. Proc Natl Acad Sci USA, 111, 12325–12330.
3. Guimer,R., Uzzi,B., Spiro,J. and Amaral,L.A.N. (2005) Team assembly mechanisms determine collaboration network structure and team performance. Science, 308, 697–702.
4. Uzzi,B., Mukherjee,S., Stringer,M. and Jones,B. (2013) Atypical Combinations and Scientific Impact. Science, 342, 468–472.
5. Latour, B., Woolgar, S (1979). Laboratory Life: The Social Construction of Scientific Facts, Beverly Hills, Sage Publications, 1979. (ISBN 0803909934) ; rd. Princeton, Princeton University Press, 1986
6. Klug,M. and Bagrow,J.P. (2016) Understanding the group dynamics and success of teams. R. Soc. open sci., 3, 160007.
7. Pentland,A. (2012) The new science of building great teams. Harvard Business Review.
8. Watts, D. (2016), The Organizational Spectroscope (`https://medium.com/@duncanjwatts/the-organizational-spectroscope-7f9f239a897c`)