Available online at www.sciencedirect.com

SciVerse ScienceDirect

Journal of Economic Theory 147 (2012) 1754–1786

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Homophily and long-run integration in social networks ☆

Yann Bramoullé [a,b], Sergio Currarini [c,d], Matthew O. Jackson [e,f,g,*], Paolo Pin [h], Brian W. Rogers [i,j]

[a] *Department of Economics, Université Laval, Canada*
[b] *Aix-Marseille School of Economics, France*
[c] *Università di Venezia, Italy*
[d] *University of Bristol, United Kingdom*
[e] *Department of Economics, Stanford University, CA, United States*
[f] *CIFAR, Canada*
[g] *The Santa Fe Institute, NM, United States*
[h] *Dipartimento di Economia Politica e Statistica, Università degli Studi di Siena, Italy*
[i] *MEDS, Kellogg School of Management, Northwestern University, IL, United States*
[j] *NICO, Northwestern University, IL, United States*

Received 1 September 2010; final version received 21 January 2012; accepted 5 February 2012

Available online 18 May 2012

Abstract

We model network formation when heterogeneous nodes enter sequentially and form connections through both random meetings and network-based search, but with type-dependent biases. We show that there is "long-run integration", whereby the composition of types in sufficiently old nodes' neighborhoods approaches the global type-distribution, provided that the network-based search is unbiased. However, younger nodes' connections still reflect the biased meetings process. We derive the type-based degree

☆ Following the suggestion of JET editors, this paper draws from two working papers developed independently: Bramoullé and Rogers (2010) [4] and Currarini, Jackson, and Pin (2010) [11]. We gratefully acknowledge financial support from the NSF under grant SES-0961481 and we thank Vincent Boucher for his research assistance. We thank the staff from the American Physical Society (APS) Journal Information Systems (http://publish.aps.org/departments/journalinformationsystems) for their assistance. We also thank Habiba Djebbari, Andrea Galeotti, Sanjeev Goyal, James Moody, Filippo Radicchi, Betsy Sinclair, Bruno Strulovici, and Adrien Vigier, as well as numerous seminar participants and the associate editor and anonymous referees for helpful comments and suggestions.

* Corresponding author at: Department of Economics, Stanford University, CA, United States.
E-mail addresses: s.currarini@unive.it (S. Currarini), jacksonm@stanford.edu (M.O. Jackson), paolo.pin@unisi.it (P. Pin), b-rogers@northwestern.edu (B.W. Rogers).
URL: http://www.stanford.edu/~jacksonm/ (M.O. Jackson).

0022-0531/$ – see front matter
http://dx.doi.org/10.1016/j.jet.2012.05.007

distributions and group-level homophily patterns when there are two types and location-based biases. Finally, we illustrate aspects of the model with an empirical application to data on citations in physics journals.

JEL classification: D85; A14; Z13

Keywords: Network formation; Social networks; Homophily; Integration; Degree distribution; Citations

1. Introduction

Homophily patterns in networks have important consequences. For example, citations across literatures can affect whether, and how quickly, ideas developed in one field diffuse into another. Homophily also affects a variety of behaviors and opportunities, with impact on the welfare of individuals connected in social networks.[1] In this paper we analyze a model that provides new insights into the patterns and emergence of homophily, and we illustrate its implications with an application to a network of scientific citations.

Our main objective is to study how homophily patterns behave in an evolving network. Do nodes become more integrated or more segregated as they age? How does this evolution depend on the link formation process? In particular, does the network become more integrated if new connections are formed at random or if they are formed *through* the existing network?

To answer these questions, we study a stochastic model of network formation in which nodes come in different types and types, in turn, affect the formation of links. We accomplish this by introducing individual heterogeneity to the framework of Jackson and Rogers [19], allowing us to focus on the issue of homophily generated through specific biases in link formation. A new node is born at each time period and forms links with existing nodes. The newborn node connects to older nodes in two ways. First, she meets nodes according to a random, but potentially type-biased, process. Second, the newborn node meets neighbors of the randomly met nodes ("friends of friends"). This is referred to as the search process and can also reflect type biases. To illustrate, consider citation patterns. Typically when writing a new paper, some references are known or found by chance by the authors while others are found *because* they are cited in known papers. Biases arise because papers may cite references with greater frequency within their own field. We examine the long-run properties of this model and the structure of the emerging network.

The biases could arise from agents' preferences over the types of their neighbors and/or from biased meeting opportunities that agents face in connecting to each other. So, in one direction we enrich a growing network model by allowing for types and biases in connections, and in another direction we bypass explicit strategic considerations by studying a process with exogenous behavioral rules. Since in the model search goes through out-links only, strategic considerations are to some extent inherently limited, since a node cannot directly increase the probability of being found through its choice of out-neighborhood. While this may not be a good assumption in some contexts, such as business partnerships or job contacts, where search presumably goes both directions along a link, it is appropriate in other contexts, such as scientific citations where the time order of publications strictly determines the direction of search.

We wish to understand the conditions under which the network becomes increasingly "integrated" over time. We consider three different notions of integration. Under *weak integration*,

[1] See [15,17,21] for more background and discussion.