Differential Network Effects on Economic Outcomes: A Structural Perspective

Eaman Jahani[1(✉)], Guillaume Saint-Jacques[2], Pål Sundsøy[3], Johannes Bjelland[3], Esteban Moro[4,5], and Alex 'Sandy' Pentland[1,5]

[1] Institute for Data, Systems and Society, MIT, Cambridge, USA
eaman@mit.edu
[2] Sloan School of Management, MIT, Cambridge, USA
[3] Telenor Group Research, Fornebu, Norway
[4] Universidad Carlos III de Madrid, Madrid, Spain
[5] Media Lab, MIT, Cambridge, USA

Abstract. In a study of about 33,000 individuals in a south Asian country, we find that structural diversity, measured as the fraction of open triads in an ego-network, shows a relatively strong association with individual income. After including all the relevant control variables, the effect of structural diversity becomes exclusive to the highly educated individuals. We hypothesize these results are due to concentrated distribution of economic opportunities among the highly educated social strata combined with homophily among members of the same group. This process leads to two important societal consequences: extra network advantages for the highly educated, similar to the rich club effect, and inadequate diffusion of economic opportunities to the low educated social strata.

Keywords: Ego-networks · Structural diversity · Income · Social status

1 Introduction

The impact of social networks on individual performance has been the subject of much interest in sociology and economics. The role of social ties on outcomes has been studied in various contexts such as health [4], education [7], productivity in firms [18,19], knowledge transfer [20] and regional prosperity [5]. Within the sociology literature, most of the attention has centered on the effect of informal social networks on the economic outcomes, and in particular job search. The foundational work by Granovetter [9,10] demonstrated that economic activity, and in particular job search, is embedded in informal social networks. Therefore, the local network influences the access to high quality employment opportunities. Later studies have built more context to the original theory of Granovetter and shown how use of social networks affects youth unemployment [1], varies by the category of the job right after college [16], interacts with education and

The original version of this chapter was revised: An acknowledgement has been added. The erratum to this chapter is available at https://doi.org/10.1007/978-3-319-67256-4_44

G.L. Ciampaglia et al. (Eds.): SocInfo 2017, Part II, LNCS 10540, pp. 41–50, 2017.
DOI: 10.1007/978-3-319-67256-4_5

leads to different outcomes depending on the extent of social isolation [6]. Nevertheless, there have been few studies on the relationship between *local network* characteristics and the most important economic outcome, namely *income*.

The closest outcome variable to income appears in the work of Lin et al. [13], which demonstrates that the social status of local contacts in the informal social network has a strong impact on the prestige of the attained job. However the prestige of the attained job as measured by Blau-Duncans SEI score [2] is too coarse since it does not capture the variations within a single occupation. A recent paper [15] did investigate the link between income and position in the global network and found that centrality of individuals is highly correlated with personal economic status. While such macro-level measures are excellent for prediction, they are too coarse for studying the flow of economic opportunities from local contacts, hence provide little sociological insight on the effect of personal social choices in the socio-economical status of individuals. Therefore in this work, we directly measure the connection between the capacity of an *individual's ego-network in terms of access to diverse information sources among the local contacts* and their *income*. This allows us to account for all the observable variation in the economic outcomes at an individual level, hence generalizing the results of [13]. Furthermore by focusing our attention on local contacts, we are able to develop insights into local network processes that provide access to economic opportunities and how the efficiency of these processes interact with other variables, such as social status.

As noted by [11], the effect of informal ego-networks on economic outcomes, and in particular job search, can be explained by four mechanisms of employer, worker, and most importantly *contact and relational heterogeneity*. Most of the previous studies on the effect of networks on economic outcomes have focused on contact heterogeneity: the variation in endowments or the micro characteristics of contacts in the network, such as their education or gender, as different manifestations of social capital. For example, [17] looked at the number of unique occupations and the proportion of white males present among the contacts and its effect of job leads. Similarly, Elliott [6] investigates how race and neighborhood location and the strength of a tie determine the level of social isolation and consequently how insulated an individual is from the labor market. Lin, Vaughn and Ensel [14] look at outcomes in job referrals and report that the occupational status of the contact, as a measure of social resources, has a strong impact on the prestige of the obtained job.

The *relational information on contacts* in an ego-network is more general than information on contact characteristics, since relational variation depends on the overall structure of the ego-network, and not only the characteristics of individual contacts. The theoretical underpinning for the impact of relational variation on economic outcomes revolves around the Granovetter weak tie theory [8]. The strong ties are associated with dense networks and triadic closures and as a result exhibit high levels of information redundancy. In contrast, weak ties tend to be bridges to diverse communities, hence have superior information novelty. With the exception of work by Burt [3] which investigated managerial success, there has not been any studies that examine the link between the full structure of the ego-network and economic outcomes, and as in our case income. Most works on the effect of relational heterogeneity have instead focused on the strength of the ties to information sources. In this study, we examine the effect of overall ego-network structure, namely its *structural diversity*, as a measure of relational heterogeneity on economic outcomes using about 33,000 surveyed individuals, a much larger scale compared to previous studies.

We have three empirical contributions in this paper. First, we examine the effect of the structural diversity of ego-networks on economic outcomes. Structural diversity is a relational characteristic that is an indication of the level of information novelty among the contacts based on their connections to each other. Second, we use income as our measure of economic outcome instead of the prestige of the jobs obtained through informal referral. By using income rather than job prestige, we generalize previous findings on network effects which have been mainly limited to labor-market outcomes. Furthermore, we believe income is an explicit measure of economic well-being which can be directly used in policy analysis. Finally, we provide evidence for the differential effects of structural diversity across different educational levels. We show that individuals with high educational status receive larger benefits from the same level of structural diversity when compared to individuals with low educational status. This result is most similar to findings in [13, 14] in which the status of the informal contact had a direct and large effect on the status of the attained job. When considered along with homophily and stratification across social status, the results of [13, 14] suggest that high status individuals receive larger benefits from their social contacts than low status individuals. This observation is in agreement with our findings. However there is an important difference between the observation of Lin et al and our findings, since in their case the differential effects are due to heterogeneity in contact characteristics. In contrast, we report the same phenomena from a relational perspective: *high status individuals have differential advantages stemming from the structure of their ego-networks, regardless of the characteristics of their contacts.*

2 Data

We use an anonymized mobile phone dataset containing one month of standard metadata in a developing country in South Asia. Our goal is to study the relationship between individual income and local structural characteristics of the network. In particular we focus on a local view of the network called ego-network. The focal node of interest is called the ego whereas all ego's connections are called alters. In addition to ego-alter edges, the ego-network includes all the edges between the alters, thus enabling us to study structural factors that are not directly controlled by the ego.

2.1 Income Data

The income categories for a random selection of more than 270,000 individuals across the country were obtained through three sequential large-scale market research household surveys. 101,500 of these surveyed individuals were customers of our phone carrier. Out of these initial 101,500 individual surveys, we restricted our data to those who are employed (no students, housewives, unemployed or retired) and are at least 25 years old. Furthermore, to prevent our results from getting biased by inactive egos without enough communication data, we limited our data only to those individuals who had a phone communication with more than 5 unique individuals over the one month period of the phone data (Approximately 20% of individuals have $degree \leq 5$). This smaller subset of surveys accounted for 32,870 subscribers who we treated as the egos in our analysis. Information about income was directly asked from the respondents, who were requested to place themselves within pre-defined income bins. Several other demographic characteristics such as education, gender, age and occupation were obtained through the same survey. Survey participants were distributed across 220 sales territories proportional to their population so that there were overall about 400 surveyed households in each sales territory. Systematic sampling was undertaken by selecting every fourth household, starting from a randomly selected geographic reference point and direction within each sub-territory. Respondents within the household were selected via the Kish grid method [12] among those who were eligible. Eligibility was defined as individuals with their own phone, between 15 and 65 years of age. The monthly income values were coded as ordinal categories from 1–13. Table 1 summarizes the correspondence between the income categories and their actual monetary value after conversion to US dollars. Figure 1 illustrates the income distribution among our egos. The Pearson correlation between the projected average income per region based on the survey results and their actual values published in official statistics is 0.925.

2.2 Social Network Data

We used one month of raw Call Detail Records (CDR) for all carrier subscribers to construct a large-scale undirected call graph, in which two individuals are connected if there is a call between the two in both direction during the observation period. Raw CDR records for each user contain the following metadata:

1. Interactions type (SMS or Call)
2. Correspondent ID (The unique identifier of the contact)
3. Direction (Incoming or Outgoing)
4. Date and time of the interaction
5. Duration of Interaction (Only valid for calls)
6. Location of cell tower serving the subscriber (Latitude and Longitude)

Edges in the call graph are weighted by the total number of phone calls between the two individuals during the observation period. From the full call graph, we extracted individual undirected ego-networks corresponding to the surveyed individuals for whom we also have income and demographics information. It is important to note that the ego-networks only contain the reciprocal links to avoid spurious one-time contacts (e.g. telemarketing) to influence our results.

Table 1. Survey relationship between household income categories and corresponding range in US dollars.

Income Category	Monthly Household Income ($)	Frequency
1	0-33	1895
2	33-78	9351
3	78-130	29718
4	130-195	28532
5	195-260	17841
6	260-325	9995
7	325-390	4536
8	390-455	3752
9	455-520	2341
10	520-585	929
11	585-651	999
12	651-1301	966
13	1301+	274

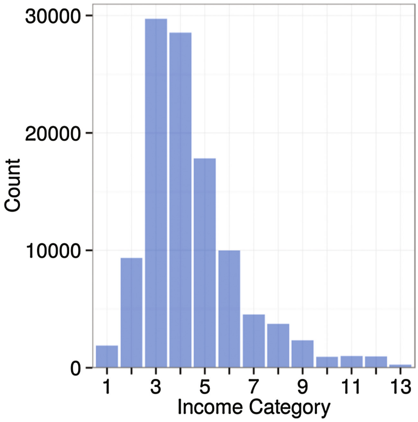

Fig. 1. Income Distribution of Egos

3 Variables

Dependent Variable: As mentioned in Sect. 2, we obtained income data for a subset of about 33,000 egos through surveys. Since income is not observed as a continuous variable, we use the middle income value in each category as representing its actual income value. As confirmed in Fig. 1, the raw income values exhibits a log-normal distribution. Therefore, middle income value of the category in USD converted to the log-scale will serve as our dependent variable.

Independent Variable: Structural diversity serves as our main independent variable. To ensure our results are robust, we perform the analysis using three different operatinalizations of structural diversity:

1. **Density** measures the completeness of the local network, and is defined as the fraction of ties from a fully-connected network that exist in the ego network. Sparsity in the ego network is an indication of structural holes and that the ego acts as a bridge between the alters, who belong to different communities. Low density also means that there is little redundancy in the ego network and most alters act as novel sources of information. Since lower values of density correspond to sparsity, we use (1-density) as our first measure of structural diversity.
2. **Clustering Coefficient** measures the fraction of closed triads in the ego network. Similar to the explanation above for density, lower values of clustering coefficient indicate diversity, non-redundancy and independence

among the contacts. Therefore, we use (1 - clustering coefficient) as our second measure of structural diversity. This would effectively measure the fraction of alter pairs between whom ego acts as a bridge which indicates the extent to which ego acts a information broker in their network.

3. **Weighted Structural Novelty** also measures the extent to which the alters are diverse and act as independent sources of information, with the main difference that it utilizes the weights of the edges. Following an argument similar to Burt [3], we compute the fraction of novelty among alters of an ego i as:

$$M_i = \frac{\sum_{j \in N(i)} (1 - \sum_{q \in N(j) \cap N(i)} p_{iq} p_{qj})}{|N(i)|} \quad i \neq j \neq q \tag{1}$$

where $N(k)$ denotes the set of k neighbors, p_{ij} is the proportion of ego i's time and energy invested in the tie with contact j:

$$p_{ij} = \frac{z_{ij}}{\sum_{q \in N(i)} z_{iq}} \quad i \neq j \tag{2}$$

where z_{ij} denote the strength of the tie or the edge weight (number of phone calls) between i and j. The term $\sum_{q \in N(j) \cap N(i)} p_{iq} p_{qj}$ in Eq. 1 represents redundancy or the extent to which information held by alter j can reach ego i through other pathways.

In addition to structural diversity, we use the level of education to demonstrate the differential effect of diversity across different social strata. The country of our interest experiences an excessive level of hereditary stratification and for this reason we believe education serves as a sufficient proxy for social status. Education will be coded as a binary variable, with high corresponding to high school, Bachelors or Masters and low corresponding to illiterate, primary school or middle school.

Control Variables: In order to control for possible confounders with income, we will include profession (a categorical variable), gender, level of education (a binary variable), age (an interval variable) and the home location of each ego on a 5×5 grid over the country (a categorical variable) as control variables in our regression analysis. Including age ensures we compare income values along the same career phases and allows us to control for long-term changes in communication patterns that are associated with variation in income. By controlling for location fixed effects, we obtain a more justified comparison of income opportunities between individuals across vastly different geographical areas (e.g. urban vs. rural). The log degree of the ego must be present as another control variable in the model, because various measures of structural diversity (e.g. density) are correlated with degree and have different scales or reasonable ranges as the ego network grows larger. For example, as the degree of the ego increases, a fully connected ego-network, corresponding to a density of 1, becomes more unlikely since the edges between alters are not independent and in particular depend on the size of the ego-network. The clustering coefficient and log degree have a correlation of 0.4 in our data.

4 Results and Discussion

Our goal is to study the effect of informal networks on career success measured in terms of income, therefore before performing the regression analysis, we excluded those egos who do not hold a valid occupation (student or housewife or retired or unemployed).

Model Specification: The equation below demonstrates the model we will use in the regression analysis:

$$Y = \beta_0 + \beta_1 * I_{high} + \beta_2 * SD + \beta_3 * SD * I_{high} + C + \epsilon \tag{3}$$

where Y and SD corresponds to income and structural diversity respectively and I_{high} is an indicator variable taking a value of 1 when ego belongs to the high education group and 0 otherwise and C corresponds to the control variables. We have three main hypothesis:

1. More structurally diverse networks are associated with higher income: $\beta_2 > 0$
2. Everything being equal, individuals with low education level have a *deficit* in their economic outcomes: $\beta_1 > 0$.
3. Individuals from a high education level have a larger *return* to the structural diversity of their networks: $\beta_3 > 0$. Note that the *return* refers to the marginal effect of structural diversity.

Table 2 shows our regression results where structural diversity is measured as clustering coefficient and in each column we successively add more control variables. We make two main observations from the results. First, all three hypothesis are validated in all models, with the exception of hypothesis 1 in model 6. Second, in model 6 which includes all the control variables, structural diversity provides no return on income for the group with low education. Effectively, only individuals with high education benefit from access to structurally diverse sources of information. We obtained similar results using the other two operationalizations of structural diversity, which points to the robustness of these observed effects. It should be noted that females in our data tend to have higher income than men; because while women in our country of study are generally housewives, those women who are employed disproportionately hold better paying jobs such as teacher or government worker.

These results confirm our main argument that information and economic opportunities are not distributed uniformly across the social network and high status individuals have an advantage in terms of their returns to networking. A potential mechanism that explains the differential returns to structural diversity relies on homophily. When the concentration of information about economic opportunities within high status and influential social strata is combined with strongly homophilous ties among individuals from the same social strata, the result is differential benefits of high status individuals from networking.

To strengthen the external validity of our findings, we plan to replicate this study on a similar data set from a different country as future work. Finally, we

Table 2. Full Regression Results. Each column successively adds more controls variables to the model. High Education and Gender are binary indicator variables. Age, Profession and Location are all categorical variables and not shown among the control variables, but their corresponding rows indicate in which models they are included. Structural diversity is measured by clustering coefficient, but the results for other operationalizations of structural diversity are similar. Degree exhibits a power law distribution, thus it is transformed to log scale. Both Structural diversity and degree are standardized.

	Log Income					
	(1)	(2)	(3)	(4)	(5)	(6)
Structural Diversity	0.037^{***}	0.019^{***}	0.019^{***}	0.019^{***}	0.008^{**}	-0.003
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Structural Diversity: High Education	0.037^{***}	0.033^{***}	0.035^{***}	0.034^{***}	0.033^{***}	0.025^{***}
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.006)
High Education	0.500^{***}	0.489^{***}	0.482^{***}	0.487^{***}	0.305^{***}	0.291^{***}
	(0.007)	(0.007)	(0.007)	(0.007)	(0.009)	(0.008)
Degree		0.047^{***}	0.048^{***}	0.048^{***}	0.033^{***}	0.036^{***}
		(0.004)	(0.004)	(0.004)	(0.003)	(0.003)
Gender Female			0.081^{***}	0.094^{***}	0.095^{***}	0.072^{***}
			(0.012)	(0.012)	(0.012)	(0.011)
Age Included	No	No	No	Yes	Yes	Yes
Profession Included	No	No	No	No	Yes	Yes
Location Included	No	No	No	No	No	Yes
Number of Variables	4	5	6	13	29	44
Observations	32,870	32,870	32,870	32,870	32,870	32,870
R^2	0.153	0.158	0.159	0.168	0.243	0.301
Adjusted R^2	0.153	0.158	0.159	0.167	0.242	0.300
Residual Std. Error	0.583	0.581	0.581	0.578	0.551	0.530
F Statistic	$1{,}980.9^{***}$	$1{,}537.8^{***}$	$1{,}241.0^{***}$	551.9^{***}	376.2^{***}	329.3^{***}

Note: $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

should mention that our claims are in no way causal since for establishing the causality we need an appropriate instrumental variable in place of structural diversity and social status. Nevertheless, we believe the observation of such differential effects, matching our theoretical expectation, renders the possibility of confounding effects unlikely.

5 Conclusion

In this study, we define the concept of *structural diversity as the extent to which the structure of a local network lacks information redundancy* and displays potential for access to diverse and independent sources of economic opportunity. We show that structural diversity, measured as the fraction of open triads in an ego-network, has a significant association with economic outcomes, mainly income, even after controlling for education, occupation, age and gender. We do this by using the ego-networks of about 33,000 individuals derived from mobile phone communication meta data matched with income and demographic variables at the individual level collected through surveys. Our findings suggest that structural diversity generally has a positive effect on income, but the benefits of structural diversity are larger for individuals with high education. This result is in agreement with a previous related study [13] in which the social status of the informal contacts determined the prestige of the obtained job. We believe this phenomenon is due to two factors: homophily and the concentrated distribution of economic opportunities among the highly educated social strata. A negative consequence of this process is the insufficient diffusion of economic opportunities to the low educated social strata.

Acknowledgement. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1122374. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Åberg, Y., Hedström, P.: Youth unemployment : a self-reinforcing process?. In: Demeulenaere, P. (ed.) Analytical Sociology and Social Mechanisms, p. 201. Cambridge University Press, New York (2011)
2. Blau, P.M., Duncan, O.D.: The American Occupational Structure, vol. 33 (1968)
3. Burt, R.S.: Structural holes and good ideas. Am. J. Sociol. **110**(2), 349–399 (2004)
4. Christakis, N., Fowler, J.: The collective dynamics of smoking in a large social network. N. Engl. J. Med. **21358**(22), 2249–2258 (2007)
5. Eagle, N., Macy, M., Claxton, R.: Network diversity and economic development. Science **335**, 1215–1220 (2012)
6. Elliott, J.R.: Social isolation and labor market insulation: network and neighborhood effects on less-educated urban workers. Sociol. Q. **40**(2), 199–216 (1999)
7. Epple, D., Romano, R.E.: Peer effects in education: a survey of the theory and evidence. Handb. Soc. Econ. **1**(1B), 1053–1163 (2011)
8. Granovetter, M.: The strength of weak ties. J. Sociol. **78**(6), 1360–1380 (1973)
9. Granovetter, M.: Economic action and social structure: the problem of embeddedness. Am. J. Sociol. **91**(3), 481–510 (1985)
10. Granovetter, M.S.: Getting a Job: A Study of Contacts and Careers, vol. 25. University of Chicago Press, Chicago (1996)
11. Ioannides, Y.M., Datcher, L.: Job information networks, neighborhood effects, and inequality. J. Econ. Lit. **42**(4), 1056–1093 (2004)
12. Kish, L.: A procedure for objective respondent selection within the household. J. Am. Stat. Assoc. **44**(247), 380 (1949)
13. Lin, N., Ensel, W.M., Vaughn, J.C.: Social resources and strength of ties : structural factors in occupational status attainment. Am. Sociol. Rev. **46**(4), 393–405 (1981)

14. Lin, N., Ensel, W.M., Vaughn, J.C.: Social resources and occupational status attainment. Am. Sociol. Rev. **59**(46), 393–405 (1981)
15. Luo, S., Morone, F., Sarraute, C., Travizano, M., Makse, H.A.: Inferring personal economic status from social network location. Nat. Commun. **8** (2017). 15227
16. Marmaros, D., Sacerdote, B.: Peer and social networks in job search. Eur. Econ. Rev. **46**(4–5), 870–879 (2002)
17. McDonald, S., Lin, N., Ao, D.: Networks of opportunity: gender, race, and job leads. Soc. Probl. **56**(3), 385–402 (2009)
18. Reagans, R., Mcevily, B.: Source network structure and knowledge transfer: the effects of cohesion and range. Adm. Sci. Q. **48**(2), 240–267 (2012)
19. Reagans, R., Zuckerman, E.W.: Networks, diversity, and productivity: the social capital of corporate R&D teams. Organ. Sci. **12**(4), 502–517 (2001)
20. Wu, L., Waber, B., Brynjolfsson, E., Pentland, A.S.: Mining face-to-face interaction networks using socimetric badges: predicting productivity in an ITC configuation task. In: ICIS. pp. 1–19 (2008)