Network diversity and economic development: a comment

Bruggeman, J.

Network Diversity and Economic Development: a Comment

Jeroen Bruggeman[*]

Abstract

Network diversity yields context-dependent benefits that are not yet fully-understood. I elaborate on a recently introduced [1] distinction between tie strength diversity and information source diversity, and explain when, how, and why they matter. The issue whether there are benefits to specialization is the key.

1   Introduction

New ideas are created by (re)combining existing ideas and applications [2, 3, 4]. Business opportunities and jobs are found amidst heterogeneous offers and demands [5]. Both novelty and economic welfare depend on information diversity, be it different kinds of information for different kinds of opportunities [6]. Seen from a network perspective, and without much knowledge about the content of information sources, it's a challenge to model diversity such that economic, scientific, artistic, and other kinds of success can be predicted.

In a recent paper in *Science*, Eagle, Macy and Claxton [1], EMC for short, found support for the relation between diversity and economic well-being in a network study of British communities. They had almost complete telephone data over a month in 2005, obviously stripped of content. Interestingly, the nodes in this network were the communities, as sources and recipients of information, not individuals, for whom numerous benefits of diversity had already been shown in other studies [7]. However, EMC's measure did not indicate diversity of sources, but diversity of time (volume of calls) spent on any given number of sources instead. This choice seems puzzling at first sight, and is not explained in their paper. I will go into their measure in some detail, and then proceed with network diversity in general.

1

2   Tie strength diversity

In the normalized Shannon entropy measure EMC propose, p_{ij} is the proportional strength, or value, of the tie (*arc*) between focal node i and contact node j, such that $\sum_{i=1}^{i=k} p_{ij} = 1$, and k_i is the number of i's contacts (*degree*). In their study, p_{ij} is community i's proportional volume of calls to community j. Although in general, $p_{ij} \neq p_{ji}$, in phone conversations and in many other social relations, information goes in both directions. Relevant exceptions are written sources of information, that can be cited but not influenced by their readers. **Normalized entropy** is defined as

$$D(i) = \frac{-\sum_{j=1}^{k} p_{ij} log(p_{ij})}{log(k_i)} \tag{1}$$

An index of economic welfare did correlate with more equally divided attention across sources as indicated by Eq.1 ($r = 0.73$).[1] This is intriguing, but we want comprehension, not just correlation. Only in the extreme case of spending almost all time on one source and almost neglecting others it's obvious that diversity of time spent reduces diversity of information. Otherwise, and net of institutions and cognitive limitations, having more sources is better, at least according to Ron Burt's theory of brokerage [5] on which EMC build (see Scott Page's additional arguments [6]). For sources to provide diverse information and opportunities indeed, they should not be connected to each other directly, and not be connected indirectly other than via the focal node itself [5]; see Fig.1. Neither of these effects is represented in EMC's measure, though, whereas other measure that they used suggested that numbers of sources ($r = 0.44$) and their lacking direct links (Burt's brokerage, $r = 0.72$) are important. (Burt's measure incorporates both effects.)

The network approach helps us to make parsimonious theories, that abstract away as much as possible from the content of the ties to make predictions as general as possible. Content matters, obviously, and a balance has to be found. It seems that to comprehend EMC's findings on the diversity of tie strength, we have to take into account some broad characteristic of their content. To this end it might help if we contrast generic phone conversations between residential communities with information transmission in the fields of research and innovation, mostly not by phone. In a comparable network, nodes are then scientific communities or technology domains [8]. There, individual inventors must have a skillful command of their sources, for example

---

[1]Eq.1 was also used to measure diversity across geographic areas, which correlated with economic welfare as well ($r = 0.58$).
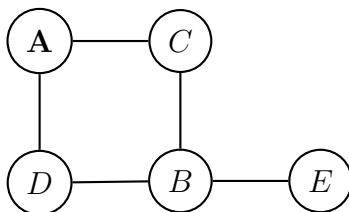
Figure 1: If all nodes divide their attention equally among their contacts, focal node A should have a lower score on diversity than node B, as in Burt's measure for brokerage; EMC's scores (eq.1) are the same for both. Furthermore, A's score should be lower if C and D were connected directly and increase redundancy rather than diversity for A, which is expressed by Burt's measure whereas EMC's scores stay equal. Finally, B exchanging information with C and D reduces opportunities for A, which is unnoticed by both EMC's and Burt's measures.

scientific literature, patents, or experts, which takes much more time and effort than maintaining business relations or asking about jobs on the phone. To cross-fertilize sophisticated knowledge successfully, knowledge brokerage must be preceded, and followed upon, by a phase of specialization in these sources [9]. An innovation-dedicated community can also self-specialize, indicated by a strong tie from the node to it self that summarizes a myriad of individuals collaborating with, or citing, each other. EMC's measure should therefore incorporate reflexive ties as well, *i.e.* allowing for the index in Eq.1 the case $i = j$. Specialization thus can happen in multiple ways, that have in common an accumulation of more densely interrelated knowledge wherein shortcuts and workarounds are discovered. Diversity of tie strength in cross-sectional data reflects combinations or alterations of specialization and brokerage, which, co-depending on network dynamics (see [9]), indicates good fortune rather than misfortune.

The content of British phone conversations we do not know, but it's clear that innovations are by far outnumbered by more mundane exchanges of information. To transfer complex information, strong ties are necessary [10], as they are for specialization, while for most interactions in daily life, like searching a job or selling an item, weak ties will do [7]. For all those more common cases, strong ties indicate redundancy rather than progressive knowledge refinement. Dedicating much attention to a few sources has therefore no advantages, or only briefly, while it precludes people and their communities from getting non-redundant information elsewhere. It seems that this explains what EMC found. We may thus conclude that their measure is very useful indeed, and focuses on one important aspect of diversity that

was previously not studied separately.[2]

3   Source diversity

To predict opportunities to create and trade, we should also come to terms with source diversity. As said, the optimal situation for a focal node is to have as many different sources as possible, for as far cognition enables and institutions allow. Our challenge is to appropriately deal with direct and indirect links between these sources.

For focal node A in Fig.1, if nodes C and D exchange information directly, A receives more redundant and less diverse information than if C and D are unconnected. Consequentially, chances for A to recombine information from them to create new opportunities decrease, be it for business, innovations, or other. Moreover, C and D no longer need A (or other nodes like B) for them to communicate; A is then out-competed with respect to benefits resulting from combinations of, and transactions between, C and D. In the case of direct links between sources, reduced diversity and increased competition are two sides of the same coin. Empirically they differ; diversity of information and other resources can in principle be observed in social interactions, while competition—if nodes do not show direct rivalry in their behavior—can't be observed but has to be inferred, from performance reduced by it.

Burt's measure does a good job at capturing the effects of both tie strength diversity and direct links between sources in one stroke. But it correlates slightly weaker with economic success than normalized entropy does [1] and it overlooks nodes one removed from the focal node that draw information or other resources from the same sources as the focal node does. If we look again at focal node A in Fig.1, B is a case in point. Suppose B provides information to C and D that is useful to them, then they have the advantage first, while A still waits or never hears about it. If, on the other hand, B uses information from C and D, the ideas B produces will be more similar to A's than if B would use sources unrelated to A. B does not necessarily reduce A's diversity but it reduces A's chances for novelty. In a rare email network study where also the content of the messages was known to the researchers, the effect of indirect links (structural equivalence)

---

[2]Some credit should go to James Coleman [11], who presented a measure of entropy (different from EMC's) in his well-known Introduction to Mathematical Sociology. Ron Burt used that measure for tie strength diversity, not to assess knowledge source specialization, though, but to show that women in an organization got early promotion if they had one strong tie to a "sponsor," a higher manager in the organization other than their own [12]—status specialization, for short.

on information diversity was indeed insignificant [13]. Other studies (based on patent data) showed that the effect of this competition on performance (citation impact) was significantly negative, though [14, 15]. There, competition was not for information itself, which does not deplete with usage [16], but for the novelty that could be created with it and valued by others. What we should measure is not diversity per se, but potentially useful diversity, to which as few as possible competitors have access. In sum, like direct links between sources discussed above, also indirect links increase competition for a focal node.

Betweenness centrality [17] is the simplest measure that captures the number of sources under both constraints, of their lacking direct links and indirect links. To broker diverse information, a focal node should sit astride on multiple *paths* (concatenations of ties) between places where "useful bits of information are likely to air, and provide a reliable flow of information to and from those places" [5]. The basic intuition dates back from 1948 [18], and its formalization came independently in 1971 [19] and in 1977 [17]. Due to its simplicity, betweenness is better comprehensible (after working through a few examples), communicable, and applicable than more sophisticated measures, of which it's generally not understood what the underlying social mechanisms would be.

For betweenness, of all paths through a focal node from here to there, only the shortest paths count. But shortest paths can still be long. Unchangeable information, like chain letters, can travel long distances [20], but response times to information are heterogeneously distributed, and the "fat tail" of slow responders strongly slows down diffusion processes [21]. In our case, shortest paths may not be short enough, because strategically relevant and manipulable information is much less reliable over longer paths, and before it reaches a focal node it has probably already been used by another middle(wo)man along its way. Diversity is by far the most useful where—and when—the news breaks [5], while "second hand brokerage" is not [22]. Moreover, long network paths strongly affect a node's betweenness scores, whereas they rarely matter for brokerage. We should therefore constrain betweenness to paths shorter than or equal to three ties in a row, and call it 3-betweenness for short. It thereby fits squarely into Fowler and Christakis' [23] "three degrees rule," a stylized fact that various sorts of social influence do not reach further than path lengths of three.

In Fig.1, B has exclusive access to E, and is the gatekeeper with the power to speed up, interrupt, or distort information from or to E [24, 25]; B thus enjoys the full benefit of paths from E to other nodes. Our focal node A, in contrast, has no exclusive sources. There is one

(shortest) path through it from C to D and another path from C to D of equal length that does not pass through A. Without any further information about the network, our initial best bet is that roughly half of the information exchange between C to D will pass through A. Assortment ("homophily" [26]), sympathy, and other factors may bias one channel in favor of another, but the associated tie strength diversity around the focal we have already captured by EMC's measure. We can complement the latter with betweenness that focuses on a different aspect of diversity. From a focal node's point of view, diversity of tie strength (or anything else) further afield is less relevant, so there we may trade off realism for parsimony. This is what betweenness does, by abstracting away from tie strength. For the presence or absence of ties, a threshold value should be established depending on the field of application. Below the threshold, information transfer is insignificantly weak and then ignored.

Generalizing these intuitions about exclusive and shared access, **3-betweenness** of focal node i is the ratio of the shortest paths, g_{jil}, from j through i to l (under the distance constraint discussed above), to all shortest paths between these two nodes, $g_{j.l}$, and then summed for all pairs of nodes in the network.[3] Formally,

$$C_B(i) = \sum_j \sum_{j<l} \frac{g_{jil}}{g_{j.l}} \qquad j \neq i \neq l, \ \ d(j,l) \leq 3. \qquad (2)$$

The reader may verify that if the number of direct or indirect links between i's sources increases, its 3-betweenness score decreases, and that direct links have a stronger impact than indirect links have.

4 Test

I tested the two measures on a network of "invisible colleges" of US inventors ($n = 417$), analogous to the British communities of citizens. In this case the ties consist of patent citations, that represent knowledge flows [27], for which I used all patents in the USA (about two million) over the period 1975—1999. The administrative units corresponding to the colleges of inventors are technology domains, wherein patents are categorized. Performance is here measured as citation impact (number of citations) over the entire period. Domains'

---

[3]Notice that if ties are (strictly) asymmetric, a path in one direction is not necessarily the same as a path in the opposite direction. Alternatives to 3-betweenness are 2-betweenness, and 4, 5, etcetera, -betweenness. It remains an empirical question if 3-betweenness predicts best. In R's igraph package, 3-betweenness for a graph G can be computed by `betweenness.estimate(G, cutoff=3)`

self-specialization is a prominent knowledge strategy; on average a domain has 214 source domains, but $p_{ii} = 0.53$, which is much higher than it would have been in an equal division of citations over source domains (0.005). To compare this network with the British community network for the effect of diversity on performance, I simplify by leaving out network dynamics (elaborated in [9]). As the average path length is short (1.49), there is no difference between betweenness and 3-betweenness. Both correlate 0.77 with performance, whereas normalized entropy correlates -0.22. The most successful technology domains thus combine brokerage with specialization, which we can clearly see by using these two measures.[4] In Burt's measure, tie strength diversity and topological diversity are combined, and as in this case they point in opposite directions (low entropy, high brokerage), that measure correlates much lower with performance, 0.19, and is less informative. Only if they point into the same direction (high entropy, high brokerage), like in EMC's study, Burt's measure is adequate. Interestingly, though, Burt's discursive theory matches the entropy and 3-betweenness measures better than his own measure. Additional tests in a variety of fields should point out if we now have both correlation and comprehension indeed.

5 Brokerage and Specialization

To assess network diversity for economic development and other accomplishments, we may start out with the elegant and simple measure of 3-betweenness. For valued graphs we complement it with normalized entropy, that should also take reflexive ties into account, if present. Subsequently, it's important to know if the field one is about to investigate is complex for its inhabitants, such that progressive knowledge or skill refinement yields benefits for them, or is relatively simple such that we may neglect small bursts of specialization. We already know that the fields of technology and science are complex, to which we may add sport, architecture, haute cuisine, art, law, and any other field where extensive schooling or training are required. (And if we don't know, we can figure it out through the effect of normalized entropy.) In all those fields we will find individuals who spend years on specializing, and have intensive contacts with relatively few and interconnected sources of knowledge, such as teachers, books, and peers. In the special case of repeated complex tasks, like building air-

---

[4]A preliminary regression model also featured a significant ($p < 0.01$) interaction effect, suggesting that brokering and specializing *at the same time* are beneficial for collectives in complex environments.

craft, specialization follows the well known learning curve [28]. We would not want to say that all those learners are wasting their time and should only muster diversity instead. Specialization to the level of mastery makes it possible to use the acquired knowledge (partly) routinely, and also enhances individuals' as well as organizations' absorptive capacity [29]. This enables to notice valuable information amidst redundancy and noise, including brokering opportunities that laymen overlook. "Chance favors the prepared mind," as Louis Pasteur said. There is of course no guarantee whatsoever that trained specialists become good brokers, or continue to be successful specialists, and they run a risk, individually and collectively, to get stuck in a local optimum of their specialization [6]—their competency trap.

In complex fields, we should expect to see the best outcome in the long run for those individuals and collectives who oscillate between, or dynamically combine, specialization and brokerage, and not stay permanently at either strategy or some place in between [9].[5] Collectives, such as large business companies with a R&D department, may employ each strategy in a different part of their organization, and teams may broker by a composition of non-overlapping specialists [32]. As we have seen, the most successful technology domains combine brokerage, to collect diverse information, with specialization, to accumulate and integrate this information to well-exploit it.[6]

We now have the tools to measure tie strength diversity and topological diversity, know more about the underlying mechanisms, can predict when specialization matters, and tell why. Finally, we should not forget that irrespective of diversity, good sources of information are substantially more beneficial than arbitrary sources are. This holds

---

[5]The cognitive processes associated with brokerage and specialization are exploration and exploitation, respectively. The human brain has different parts for each [30], and noradrenaline helps regulating the dynamic balance between the two [31]. When the temporal aspect is overlooked, paradoxes may result. When a broker gets to know her contacts, or sources, well, she may exploit them, whereas progressive specialization is only possible through exploring more efficient shortcuts or (re)combinations. The paradoxes vanish when time is taken into account.

[6]Normalized entropy as discussed here captures both *self* and *source specialization*. As a refinement of source specialization, nodes can also exchange information dyadically, which we might call *mutualistic specialization*. For technology domains, it correlated positively with performance. A more interrelated knowledge base also results from *cluster specialization* [4], i.e. when a focal node's sources draw ideas from each other, either with or without the focal node's own doing. When a node's local clustering increases, betweenness necessarily decreases. Furthermore, there is *geographic specialization* in specific, often proximate, areas, which also holds for patent citations [33]. Finally, there is *status specialization* (footnote 2), i.e. a preference for linking to high status nodes. In the case of technology domains, it coincides with auto-regression (see main text).

throughout society, from technology domains [15] to philosophers [34]. High performing nodes thus have an important spillover, of higher quality knowledge for their network neighbors who specialize in them (an instance of network *auto-regression*). On this note, we may end with some practical advice. First, have good sources of information, and keep in mind that having some good sources is better than having just many. Second, make sure they are diverse. Third, integrate and master complex information from these sources through specialization.

References

[1] N. Eagle, M. Macy, R. Claxton, *Science* **328**, 1029 (2010).

[2] H. Poincaré, *La science et l'hypothèse* (Flammarion, Paris, 1902).

[3] M. Weitzman, *Q. J. Econ.* **113**, 331 (1998).

[4] R. Burt, *Am. J. Sociol.* **110**, 349 (2004).

[5] R. Burt, *Structural Holes* (Harvard U.P. Cambridge, MA, 1992).

[6] S. Page, *The Difference* (Princeton U.P., Princeton, NJ, 2007).

[7] M. Granovetter, *Sociol. Th.* **1**, 201 (1983).

[8] G. Davis, C. Marquis, *Organiz. Sci.* **16**, 332 (2005).

[9] G. Carnabuci, J. Bruggeman, *Soc. Forces* **88**, 607 (2009).

[10] M. Hansen, *Admin. Sci. Q.* **44**, 82 (1999).

[11] J. Coleman *Introduction to Mathematical Sociology* (Free Press, N.Y., 1964).

[12] R. Burt, *Rationality Soc.* **10**, 5 (1998).

[13] S. Aral, M. Van Alstyne, *Proc. Acad. Manag.* (2007).

[14] J. Podolny, T. Stuart, *Am. J. Sociol.* **100**, 1224 (1995).

[15] G. Carnabuci, *Soc. Forces* **88**, 2163 (2010).

[16] K. Arrow, in R. Nelson (ed.) *The Rate and Direction of Inventive Activity* (Princeton U.P., Princeton, NJ, 1962).

[17] L. Freeman, *Sociometry* **40**, 35 (1977).

[18] A. Bavelas, *Appl. Anthrop.* **7**, 16 (1948).

[19] J. Anthonisse, Report (Math. Center, Amsterdam, 1971).

[20] D. Liben-Nowell, J. Kleinberg, *PNAS* **105**, 4633 (2008).

[21] J. Iribarren, E. Moro, *Phys. Rev. E* **103**, 038702 (2009).

[22] R. Burt, *Acad. Manag. J.* **50**, 119 (2007).

[23] N. Christakis, J. Fowler, *Connected* (Little Brown, N.Y., 2009).

[24] J. Boissevain, *Friends of Friends* (Blackwell, Oxford, 1974).

[25] L. Freeman, *Qual. Quant.* **14**, 585 (1980).

[26] E. Rogers, *The Diffusion of Innovations* (Free Press, N.Y., 5th ed, 2003).

[27] R. Henderson, A. Jaffe, M. Trajtenberg. *Am. Econ. Rev.* **95**, 450 (2005).

[28] L. Argote, D. Epple, *Science* **247**, 929 (1990).

[29] D. Levinthal, W. Cohen, *Admin. Sci. Q.* **35**, 128 (1990).

[30] N. Daw, J. O'Doherty, P. Dayan, B. Seymour, R. Dolan, *Nature* **441**, 876 (2006).

[31] J. Cohen, G. Aston-Jones, *Nature* **436**, 47 (2005).

[32] R. Guimerà, B. Uzzi, J. Spiro, L. Nunes Amaral, *Science* **308**, 607 (2005).

[33] A. Jaffe, M. Trajtenberg, R. Henderson, *Q. J. Econ.* **108**, 557 (1993).

[34] R. Collins, *Phil. Soc. Sci* **30**, 157 (2000).