Global Village or Cyber-Balkans?

Modeling and Measuring the Integration of Electronic Communities

Marshall Van Alstyne
Boston University
595 Commonwealth Ave
Boston, MA 02215
mva@bu.edu (617) 358-3571

Erik Brynjolfsson
MIT Sloan School of Management
30 Wadsworth St., E53-313
Cambridge, MA 02139
erikb@mit.edu (617) 253-4319

This version: December 27, 2004.

KEYWORDS: Globalization (AF1301), Information Economy (AD02), Information Flows (DD07), Computerization of Society (BD0101), Organizational Structure (DA03), Economic Impacts (BA01), Balkanization, Social Networks

ABSTRACT: Information technology can link geographically separated people and help them locate interesting or useful resources. These attributes have the potential to bridge gaps and unite communities. Paradoxically, they also have the potential to fragment interaction and divide groups. Advances in technology can make it easier for people to spend more time on special interests and to screen out unwanted contact. Geographic boundaries can thus be supplanted by boundaries on other dimensions. This paper formally defines a precise set of measures of information integration and develops a model of individual knowledge profiles and community affiliation. These factors suggest specific conditions under which improved access, search, and screening can either integrate or fragment interaction on various dimensions. As IT capabilities continue to improve, preferences -- not geography or technology -- become the key determinants of community boundaries.

ACKNOWLEDGMENTS: This paper has benefited from helpful comments by Mark Ackerman, Linda Argote, Brian Butler, Piotr Dollar, Rob Fichman, Rob Kling, Paul Laskowski, Andrea Meyer, Dana Meyer, David Meyer, Jorge Schement, participants in the Cambridge Roundtable, the MIT Coordination Science Workshop, a seminar at the University of Michigan, the International Conference on Information Systems, and five anonymous reviewers. We thank the National Science Foundation (IIS-0085725 and 9876233) for partial financial support.

A later and more polished version is forthcoming in *Management Science*. Please refer to that version for specifics or citation.

I. Introduction -- The Emerging Global Village?

Explosive growth in computer mediated and networked communications can shrink distances and facilitate information exchange among people of various backgrounds. Telecommunications policy in the United States – and other countries – has long resolved to extend access to all levels of society, assuming that this will automatically foster greater information exchange while boosting economic growth (e.g. NTIA, 1993).

Empowered by search engines, recommender systems, search agents and automatic filters, information technology (IT) users are spending more of their waking hours on the Internet, choosing to interact with information sources customized to their individual interests. But, does the emergence of a global information infrastructure necessarily imply the emergence of the global village -- a virtual community of neighbors freed of geographic constraints? Or, will the borders merely shift from those based on geography to those based on interest?

In this paper, we show that an emerging global village represents only one of a range of possible outcomes. Improved communications access and filtering technologies can, in some circumstances, lead to more fragmented intellectual and social interaction. In particular, we show that preferences can reshape social, intellectual and economic neighborhoods as distinct from those based on geography. Just as separation in physical space can divide geographic groups, we find that separation in virtual knowledge space can divide special interest groups. In certain cases, the latter can be more insular. We introduce several formal indices of integration and then show both algebraically and graphically the conditions under which these indices rise or fall with different preferences and levels of access.

The conclusion that increased connectivity and improved filtering can actually lead to less integration is based on two observations. First, bounded rationality, a limit on the human capacity for information processing (Simon, 1957), can lead to specialization, which decreases the range of overlapping activities. As IT eliminates geographical constraints on interaction, the constraints of bounded rationality become increasingly important. Information transmission and bandwidth have increased across all distances except the last 12 inches -- between people and machines. Regardless of how fast data scrolls across the screen, absorption is bounded. In the limit, people must choose some information contacts over others. Filters, even sophisticated electronic filters, must be selective in order to provide value. Thus, certain contacts, ideas or both will be screened out.

The second observation is that IT can provide a lubricant that enables the satisfaction of preferences against the friction of geography. On the one hand, those with a preference for specialization, whether intrinsic or driven by external rewards, may seek more focused contact than available locally. Thus, local heterogeneity can give way to virtual homogeneity as specialized communities coalesce across geographic boundaries. On the other hand, preferences for broader knowledge, or even randomized information, can also be indulged. In the presence of IT, a taste for diverse interaction leads to greater integration – underscoring how the technology serves mainly to amplify individual preferences. IT does not predetermine one outcome.

The same mechanisms that affect the specialization of knowledge also affect the degree to which interactions among people and communities become more or less integrated. The Internet can provide access to millions of other users and a wide range of knowledge sources, but no one can interact with all of them. Bounded rationality implies that a citizen of cyberspace still has a finite set of "neighbors" with whom he or she can meaningfully interact, but that non-geographic criteria increasingly influence the selection of these neighbors. Non-geographic criteria for selecting acquaintances can include common interests, status, economic class, academic discipline, religion, politics or ethnic group. In some cases, the result can be a greater balkanization along dimensions that matter far more than geography, while in other cases more diverse communities can emerge. Our analysis suggests that automatic search tools and filters that route communications among people based on their views, reputations, past statements or personal characteristics are not necessarily benign in their effects.

Preferences themselves need not remain unaffected by such tools. Because the Internet makes it easier to find like-minded individuals, it can facilitate the creation and strength of fringe communities that have a common ideology but are dispersed geographically. Thus, particle physicists, oenophiles, Star Trek fans, and members of militia groups have used the Internet to find each other, swap information and stoke each others' passions. In many cases, their heated dialogues might never have reached critical mass as long as geographic separation diluted them to a few participants per million. Once connected, their subsequent interactions can further polarize their views or even ignite calls-to-action (Sunstein, 2002). The Internet can also facilitate the de facto secession of individuals or groups from their geographic neighborhoods. One study found that increased hours spent using the Internet can be strongly associated with a loss of contact with one's social environment and spending less time with human beings (Nie & Erbring, 2000). Another study found that users decreased their local knowledge while their knowledge of national events remained about the same (Kraut, et. al. 2002). Consistent with the predisposition arguments presented below, the latter study also found that introverts decreased on measures of community involvement and increased in loneliness while extroverts increased their involvement and decreased in loneliness. The Internet can apparently lead to spending less time interacting with geographic neighbors, isolating individuals on some dimensions even as it integrates them on others.

We do not argue that increased specialization or balkanization must always result from increased connectivity. On the contrary, we believe that the Internet has enormous potential to elevate the nature of human interaction. Indeed, we find that if preferences favor diversity, increased connectivity reduces specialization and increases integration. Strong ties and social bonding provide important social benefits (Wellman & Wortley 1990; Putnam 2000). However, our analysis also indicates that, other factors being equal, all that is required to reduce integration in most cases is that preferred interactions are more focused than existing interactions. A desire for increased focus and improved filtering of noisy communications is a natural response to data and computational overload. Although the conventional wisdom has stressed the integrating effects of the technology, we examine critically the claim that a global village is the inexorable result of increased connectivity and develop a suite of formal measures to address this question.

II. Related Literature

To characterize group information sharing, we draw on related literature from a variety of perspectives including theories of attraction (Blau, 1977), dynamic social interaction (Latane, 1996), group stability (Carley, 1990; 91), group diversity (Ancona & Caldwell, 1992), social networks (White, et al. 1976; Wellman & Wortley 1990; Wellman & Gulia 1997) network measures (Banks & Carley, 1996; Sunil, Banks & Carley 1995; Teachman, 1980; Wasserman & Faust, 1994; Watts & Strogatz, 1998) and diffusion models (Valente 1995).

Like Blau (1977), we use an attribute vector such as age, sex, race, religion, and employment to predict social differentiation, group formation, and individual tendencies toward social interaction, but we focus on information access. His homophily model of attributes, for example, predicts that two white male postal workers share more in common than either might share with a black female executive. Based on differences among individuals and the assumption that influence declines with distance, Latane (1996) argues that group patterns emerge as a function of the strength, immediacy, and number of social factors acting on individuals. Latane's Dynamic Social Impact Theory holds that people become more similar to their neighbors, leading to spatial clustering, and that changing patterns may exhibit non-linearity as opinions resist outside pressure up to a threshold, which we model explicitly in Corollary 2.1. An empirical study in support of this theory found that group members came to resemble their neighbors in electronic space, opinions on unrelated topics became correlated, and majority factions increased in size, but minority factions became more coherent (Latane & Bourgeois, 1996).

Group stability is also considered in Carley's (1990, 1991) “constructural” model where groups “form and endure because of discrepancies in who knows what.” Shared knowledge leads to interaction and in turn interaction leads to shared knowledge. The modeling parameters and analysis resemble those introduced here, with a few exceptions. First, Carley's simulation analysis tracks the complex dynamic character of group boundaries over time. In contrast, our derivations are analytical and focus on comparative static results and equilibrium conditions. Second, most models of this type (e.g., triad completion, constructural, degree variance) eventually homogenize in the sense that interaction probabilities between all pairs of agents become equal (Banks & Carley, 1996). In our model, homogenization and balkanization can both result. The key difference is the interaction of preferences with bounded capacity; for if agents in our model had unbounded capacity, integration would always result. Indeed, even with bounded capacity and a preference for diversity, integration still results. In this sense, the models are consistent and complementary.

Unlike "learning" models in the literature, our model does not explicitly treat information spreading perfectly from person to person. Simulations have shown that results presented here are qualitatively similar if either information decays with time or attenuates with distance (as in Zipf, 1946), or is "sticky" (as in von Hippel, 1998) in terms of the expertise required to process it. Either factor can move equilibrium knowledge profiles from homogeneity towards clustering, contingent on preferences. If perfect knowledge transfers are allowed, but extreme preferences prevent inter-group interaction, then subsequent results are unchanged. If learning is allowed but balkanization refers to group formation apart from what members know, then results are also unchanged.

A contrasting perspective appears in Watts & Strogatz (1998), which models small world phenomena. Their model considers paths between agents in which groups exhibit a high degree of local clustering but also a fairly short average path length between individuals. Through simulation and analysis, they show that adding random links to a structured network, which has high local clustering and long average path lengths, can reduce average path length much more rapidly than it reduces clustering. Thus local communities could appear to have numerous in-group ties while the distance to members of out-groups appears fairly short – an idea first captured in Milgram's phrase "six degrees of separation," implying that any two people across the globe could be linked by a chain of only six people.[1]

To the extent that data diffuses more rapidly, shorter paths between distant people will promote more integrated information. Transfer also depends, however, on preferences. Intermediate people in a chain must be willing to serve as conduits for data that need not necessarily pertain to them. In a dramatic demonstration of this, Dodds et al. tried to recreate the Milgram letter passing experiment. Despite the ease of using e-mail over standard mail, fully 98% of chains failed to complete (Dodds, Muhamed, Watts, 2003).[2] Thus news of popular interest, terrorist attacks, and jokes-of-the-day, diffuse rapidly, while subtle ideas or those of parochial interest, like new mathematical theorems, diffuse only slowly. Subtle ideas may also require sophisticated knowledge to convey. Subtle information is less likely to diffuse

[1] The interested reader can explore tutorials and online simulations of all three models – balkanization, constructural, and small world – at www.IndigoSim.org.

[2] The authors write "We conclude that although global social networks are, in principle, searchable, actual success depends sensitively on individual incentives" p. 827 i.e. preferences matter.

rapidly without loss from node to node, as the child's game of "telephone" illustrates even for simple rumors. Related critical mass and threshold models of diffusion also appear in Valente (1995). One difference is that Valente allows for “opinion leaders” whereas the present research treats the agents equally in the analysis.

Information integration also differs from group integration. Although the former measures the knowledge individuals have in common, the latter measures the communities they commonly form. The first considers the overlap in what people know while the second considers the overlap in how they spend their time. As IT can affect both, we introduce measures of knowledge profiles and community membership that are built on the same basic constructs.

Existing literature provides many useful indices of network structure. These include homogeneity (Blau, 1977; Banks & Carley 1996; Carley 1991, 1995), diversity and complexity (Teachman 1980), centrality and vulnerability (Freeman, 1979; Malone, 1988; Van Alstyne, 1997), dyads, triads, and link cuts (Wasserman & Faust, 1994), tie strength, blocks, and structural holes (Granovetter, 1973; White 1976; Burt 1993), link evolution (Sanil et. al., 1995), and even integration and polarization (Kaufer & Carley, 1993; Banks & Carley 1996).

Complementing this literature, our research provides specific new measures of clustering that differ from measures of diversity and homogeneity and are analytically tractable. We also borrow from information retrieval theory and use these indices to examine theoretical implications of changing interconnectivity, searching, and screening.

III. Modeling People and Resources – Measures of Integration

To examine community interactions, we construct a model of individual contact and information resource distribution. By "electronic community" we mean groups of individuals that participate in online contact and information sharing. We also acknowledge that there are broader conceptualizations of community that include, for example, companionship and emotional support (Wellman & Gulia, 1997). Electronic communities are better able to share nonrival resources than say, goods or services (Carley, 1991). Since "balkanization" lends itself to several interpretations, we introduce multiple measures including: the overlap of contact between groups, the distance of communication, and the level of concentration in information resources. Let the agents be enumerated as $i, j \in \{1, 2, 3, \ldots \mathcal{N}\}$ where $\mathcal{N}$ is the size of the total population. Then we can say that access $\mathcal{A}$ improves as it increases from 1 to $\mathcal{N}$ and that $\mathcal{A}/\mathcal{N}$ represents the fraction of the population any given agent i can potentially reach. Also, each agent has $\mathcal{C}$ channels: the maximum number of people from the population he or she can contact simultaneously assuming bounded rationality. For example, the telephone network might grant one access to billions of people ($\mathcal{A}$) but time constraints might permit mutual interaction with no more than several dozen in a given day ($\mathcal{C}$).

Adopting the convention of an information resource as a knowledge base represented by k_{it} we can associate knowledge with individual agents i in terms of both a type $t \in \{1, 2, 3, \ldots \mathcal{T}\}$ and amount k known.[3] Importantly, this also allows us to distinguish access by type and to characterize knowledge profiles by agent. Let the knowledge profile $\mathcal{P}_i$ of agent i be a vector of how much he knows about each topic $\mathcal{P}_i = [k_{i_1}, k_{i_2}, \ldots k_{i_T}]$. Each agent can thus be mapped to a unique point in "knowledge space" which is analogous to his or her geographic location. If an agent starts with only a single type of information and has knowledge profile $\mathcal{P}_i = [0, 0, \ldots k_{it}, \ldots 0]$ then allowing access to an agent j who has knowledge of a different topic s can potentially provide agent i with a profile of $\mathcal{P}_i = [0, 0, k_{is}, k_{it}, \ldots 0]$. Then, if k_t is the total knowledge of a given type i.e., $k_t = \Sigma k_{it}$, we can describe the total knowledge existing in a population as $\mathcal{K} = [k_1, k_2, \ldots k_T]$. For simplicity, we do not require agents with the same type of knowledge to know exactly the same information. Thus agents with overlapping information can connect with a net gain in resources. Under these assumptions, increasing access has the attractive property of increasing an agent's knowledge profile towards full information where $\| \mathcal{P}_i \| / \| \mathcal{K} \| = 1$. The magnitude of the knowledge profile indicates how close an individual agent comes to accessing the full information available to a society of individuals. Using this terminology, we now have the ability to calculate several useful indices.

Shared Knowledge Index: Borrowing a measure from information retrieval theory (Manning & Schutz, 2000), define the degree of "similarity" between knowledge profiles $\mathcal{P}_i$ and $\mathcal{P}_j$ as the cosine of Θ_{ij}, the angle between them.

[3]In this paper, we model a cardinal measure of knowledge. However, for most analyses, a simple Boolean value, 0 or 1, will suffice with a commensurately larger number of knowledge types, i.e., either a fact is known or it isn't.

Definition: The "similarity," S_{ij}, between two individuals in "knowledge space" is given by: $S_{ij} = \mathrm{Cos}(\Theta_{ij}) = \mathcal{P}_i \cdot \mathcal{P}_j / \|\mathcal{P}_i\| \, \|\mathcal{P}_j\|$.

$\mathrm{Cos}(\Theta_{ij})$ approaches 1 as profiles become more similar and approaches 0 as they grow farther apart.[4] Occasionally, it will also be convenient to index a group by the overall average of their profiles.

Shared Knowledge Distance: An alternative to S_{ij} is to compute the "distance" $\mathcal{D}_{ij}$ between knowledge profiles by applying a distance metric to their difference.

Definition: The "distance" between two individuals, i and j, in "knowledge space" is:

$$D_{ij} = \left\| P_i - P_j \right\| = \sqrt{(k_{i1} - k_{j1})^2 + (k_{i2} - k_{j2})^2 + \ldots (k_{iT} - k_{jT})^2}$$

If both agents i and j have access to exactly the same knowledge bases[5], then this expression reduces to zero, but this index can otherwise take on values in $\mathbb{R}^+$.

Integrated Affiliations: Based on these profiles, we can also define an index of how much agents' group memberships overlap. An agent, who starts out with resources of type t_1, can increase affiliations by gaining access to other types t_2. For an index of affiliation, we want a measure that increases when communities overlap and that decreases with the number of separate communities. Let the members affiliated with a community of type t be given by $\mathcal{M}(t)$ so that we can derive a metric of group overlap, generalizing a covariance measure of two-way overlap (Donath, 1995).

Definition: The "index of integrated affiliation," μ_A, for a population is given by:

$$\mu_A = \frac{1}{T} \frac{1}{T-1} \sum_{t \in \{1,2 \ldots T\}} \sum_{s \neq t} \frac{\left\| M(t) \cap M(s) \right\|^2}{\left\| M(t) \right\| \left\| M(s) \right\|}$$

This index ranges from a low of 0, if every community is closed and shares no members in common with other communities, to a high of 1 if every individual is a member of every community (it is undefined if everyone is the same type). The more diverse an agent's associations, the more he or she raises the index of integrated affiliation.

Banks & Carley (1996) propose an alternative balkanization measure based on edge formation. That is, the propensity for node i to link to j depends on what other partners they hold in common. While

---

[4] Consistent with a positive distance metric, we assume no "negative" knowledge, thus $\mathrm{Cos}(\Theta_{\iota\varphi})$ is non-negative.

[5] If we wish to allow for the possibility of knowledge overlap, then type differences become set differences, e.g., $|\kappa_{\iota\tau} \cup \kappa_{\varphi\tau} - (\kappa_{\iota\tau} \cap \kappa_{\varphi\tau})|$. Similar changes in other indices provide consistent results.

this is more amenable to analysis of pairwise interaction and edge formation, it is also more difficult to interpret as a population-wide statistic. For example, the population average propensity when in-groups have high probabilities and out-groups have low probabilities may resemble that of a uniform population with middling probabilities. The current proposal complements prior work by providing a population level index based on category association in addition to edge formation.

Integrated communication: This index measures the integration of channel paths or who talks to whom. In a fragmented community, agents communicate in clusters or possibly not at all. In a fully integrated community, each agent communicates with everyone. For an index of integrated communication, we require a measure that decreases in the number of isolated agents and that increases each time agents establish a connection. If agents are connected in a graph, let the communication distance between two agents i and j be the total number of links $\mathcal{L}_{ij}$ on the shortest path between them. Note that $\mathcal{L}_{ij}$ need not equal $\mathcal{L}_{ji}$ if communication is directional or agents use different intermediaries. Also, since agents do not need to connect to themselves, the least upper bound on a chain of connections among $\mathcal{N}$ agents is $\mathcal{N}$-1. If no chain of connections exists between i and j, define the distance to be $\mathcal{N}$.

<u>*Definition*</u>: With these terms, define μ_C, the measure of integrated communication, as:

$$\mu_c = 1 - \frac{1}{N^2} \sum_{i \in \{1,2 \ldots N\}} \sum_{j \neq i} \frac{l_{ij}}{N-1}$$

Thus $\mu_C \in [0, 1)$, approaching 1 (i.e., 1-ε) as the population becomes large for highly integrated topologies. It reaches its lowest value when every agent is a single disconnected island and reaches its highest value when every agent is directly connected.

Integrated Resources: The degree to which knowledge bases are concentrated can vary independently of whether specific agents are directly connected, i.e., whether communication itself is integrated. A refusal to share, for example, would balkanize information despite the existence of a channel whereas access via an alternate source would integrate the same resource. For an index of integrated information, we require the measure to decrease as more resources become inaccessible to any single agent and also to decrease as more agents find the same resource inaccessible.

<u>*Definition*</u>: We define the index of integrated information μ_I to be

$$\mu_I = \frac{1}{N}\frac{1}{T}\sum_{i \in \{1,2\ldots N\}}\sum_{t \in \{1,2\ldots T\}}\left(\frac{k_{it}}{k_t}\right)^2$$

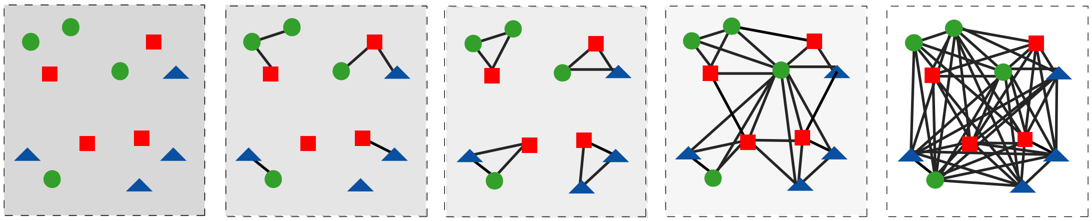
This index ranges from 0, when a single agent has exclusive access to all of a society's knowledge resources $\mathcal{K}$ to a maximum of 1 when the entire population has access to $\mathcal{K}$. Although we base this index on information shares, μ_I could equally well be used to measure other resource concentrations.[6] Measure μ_I parallels the measure of population diversity $D=1-\Sigma p^2$ with p giving the proportion of a population falling into any given category and both share the same basic properties as the entropy measure $H=-\Sigma p Log(p)$ (Teachman, 1980).

The similarity measure $S_{ij} = \text{Cos}(\Theta_{ij})$ provides an index of 'likeness' of individual access. The integration measure μ_A indexes the diversity of group interactions. The two additional measures complement these two; μ_C refers to communications and μ_I to information resources.[7] Although they can move quasi-independently, in most cases the results tend to be qualitatively similar, so we focus on S_{ij} and $\mu_{\mathcal{A}}$. This collection of indices provides a way to compare both individuals and groups within a society based on the same constructs of information access and affiliation. An example of S_{ij}, $\mu_{\mathcal{A}}$, μ_C, and μ_I is provided in the next section.

IV. Unbounding Geography

As communication costs fall generally, the cost of connecting individual agents also falls. If the costs are too high, no two agents communicate; if the costs are negligible, all agents can communicate. With IT costs falling dramatically, inter-connectivity is likely to increase (Malone & Smith, 1987; Wyner & Malone, 1996). One possible progression is a move from completely isolated agents to completely interconnected agents, as in Figure 1. We use these to illustrate the indices of integration.

[6]This measure adapts the Herfindahl index of market share concentration to multiple dimensions. This index is similar in spirit to a Gini index, which measures concentration as the sum of deviations from the 45 degree line on the plot of sorted resources versus sorted individuals. If income plots at 45˚ and is equally distributed, then the bottom 25% of the population earned 25% of all income and so on, for example.
[7]Note that indices of "balkanization," β, could just as easily be represented as $1-\mu$.



Connectivity levels increase as communication costs fall from left to right.

This example conforms to popular ideas on the emergence of networked infrastructure. When communication costs are prohibitive, these twelve agents operate in isolation with incomplete knowledge of global information, as in Figure 1.A. As communication costs fall, clusters of communication emerge, allowing agents to share information and gain a less fragmented understanding. This is shown in intermediate frames. Once costs become negligible, a fully connected community emerges, permitting everyone access to full knowledge of events, as in Figure 1.E. From left to right, knowledge profiles grow from their greatest fragmentation to their least fragmentation while community "integration" increases. Different agents, represented by different shapes, may have different information requirements or communication interests. These potential preferences will motivate subsequent observations on how much communication actually occurs. For instance, even when agents are *able* to communicate with all other agents, they may not be willing to do so. For now, we assume that agents are both willing and able, so the potential of increased access is, in fact, realized. The basic intuition follows formally below.

Proposition 1 -- Without bounded rationality constraints, global access maximizes integration when agents accept all communication. That is, $\mu_A = 1$, $\mathcal{S}_{ij} = 1$, and $\mathcal{D}_{ij} = 0$. Agents belong to the same group, their knowledge profiles are the same, and information distance is minimized.

Proof: Without bounded rationality, $\mathcal{C} \geq \mathcal{N}$, so every agent can connect to every knowledge base and $\forall\ i, j$ we have that $\mathcal{P}_i = \mathcal{P}_j = \mathcal{K}$ thus $\| \mathcal{P}_i - \mathcal{P}_j \| = 0$ implying $\text{Cos}(\Theta_{ij}) = 1$ and $\mathcal{D}_{ij} = 0$. Also, if every agent has access to all topics, then $\forall\ t$, $\mathcal{M}(t) = \{1, 2, \ldots \mathcal{N}\}$ i.e., membership is the population. Therefore $\mu_{\mathcal{A}} = 1 - (1/\mathcal{T})(1/\ \mathcal{T}-1)(\ \mathcal{T})(\ \mathcal{T}-1)(\mathcal{N}^2/\mathcal{N}^2) = 0$. □

The table below shows average similarity and integration for the graphs in Figure 1.

Index	Figure 1.A	Figure 1.B	Figure 1.C	Figure 1.D	Figure 1.E
Average($\mathcal{S}_{ij}$)	.27	.55	.77	.84	1
$\mu_{\mathcal{A}}$	0	.25	.65	.83	1
$\mu_{\mathcal{C}}$	0	.11	.17	.87	.92
$\mu_{\mathcal{I}}$	.02	.05	.08	.17	1

Average($\mathcal{D}_{ij}$)	1.02	1.34	1.15	2.29	0

Table 1 – Measures of integration change with the communication changes of Figure 1.

In this example, there are 4 agents of each type, so their knowledge profiles overlap somewhat in Figure 1.A. If there were 12 separate types, the similarity measure would be 0 in Figure 1.A. By Figure 1.E, all agents have access to society's information, so knowledge profiles are identical. Communities of types in Figure 1.A, however, share no members in common, so μ_A indicates complete segregation. Once the types are completely interconnected, this index rises to 1. The other metrics for communications distance μ_C and information concentration μ_I are provided for illustration. Note that expected average distance moves monotonically although observed average distance moves non-monotonically in small samples, being affected by larger contact networks of central agents.

The rise in integration associated with improved access in this simple model is consistent with the common view that telecommunications, and the Internet in particular, foster an emerging global village.

V. Bounding Rationality

The elimination of geographic constraints, under Proposition 1, assumes a lack of communications constraints such as bounded rationality or vetoed interaction. The physical potential for connectivity need not imply actual connectivity when either party at one end of a connection is preoccupied or otherwise unwilling to interact. Limitations on interaction exist due to (1) bounded rationality, e.g., e-mail overload[8], (2) missing or unshared vocabulary, e.g., specialized medical terminology, (3) insufficient bandwidth, e.g., even video-conferencing may provide insufficient context for first meetings, (4) unwillingness to share information despite the technological capability to do so due to inadequate incentives (see e.g., Orlikowski, 1992), lack of trust or other economic or social factors.[9]

[8]Unconstrained communication can be burdensome. During one police investigation, an Internet request for information resulted in too many false leads during a time-sensitive abduction (Leslie, 1995). Newsgroup readers actively discourage posting irrelevant material -- off-topic news, solicitations, personal attacks -- partly because of the time and nuisance costs it imposes on the community.

[9] When sender and receiver disagree on the quantity of information to be exchanged, the “short-sider” rule typically adjudicates: the quantity of information preferred by the party with the smallest preference (possibly zero) prevails.

One implication of such limitations has been the emergence of striking geographic regularities (Allen, 1984; Zipf ,1946; Abler, et. al., 1971; Latane, 1996). Whether measuring calls, travel, or mail between two cities, interactions tend to distribute linearly on log scale (Zipf, 1946).

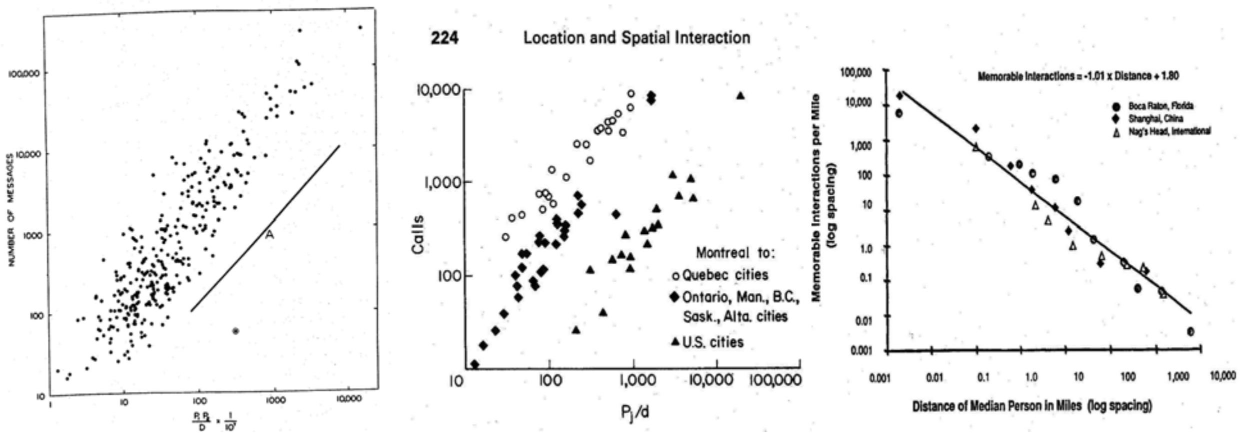


Figure 2.a, 2.b, 2.c – The first shows calls between 311 city pairs (Zipf, 1946). The second shows calls from Montreal (Abler et. al., 1980 citing McKay). The third shows recalled interactions in three different locations (Latane, 1996).

While distance plays an important role, that role appears to be decreasing (Cairncross, 1997). Coffman & Odlyzko (2002) show that the distance-sensitive element of communications costs has decreased dramatically both for voice and data and the relative proportion of long distance to local calls has consistently risen.[10]

Given the strength of geographic screening, if its effect diminishes, what emerges to take its place? A plausible alternative is to choose based on similar interests or similar attributes (Byrne,1971; Blau, 1977; Kaufer & Carley,1993; Latane, 1996). Empirical evidence also suggests that increasing socioeconomic gaps or ethnic heterogeneity reduce participation in social activity at statistically significant levels (Alessina & Ferrara, 2000). The plots of Figure 2 could be redrawn based on proximity of interests, attributes, or attitudes. Let a 'homophily' preference of a profile $\mathcal{P}_i$ be defined as the desire to communicate with another profile $\mathcal{P}_i$ agent if available, and with any agent $j\neq i$ otherwise.

For modeling purposes, we take preferences as fixed (relaxing this later) and allow agents to communicate first with their desired contact types in the neighborhood afforded by access $\mathcal{A}$. Assuming

[10] Data calls to local ISPs are an interesting exception to this trend (although one could argue that these local calls may represent much longer-distance online communications). Nonetheless, calls to human beings show strongly consistent trends toward an increase in the proportion of long distance.

that agents prefer diverse contact types to no contact, they then allocate excess capacity to diverse types as if mixing with the population at large. Note that a preference for exclusively similar contacts balkanizes initial interactions, which we analyze after the base case. The interaction of bounded rationality, access, and similar preferences yields Proposition 2.

Proposition 2 – Virtual communities decrease expected integration and increase information distance relative to geographic communities given bounded rationality, $\mathcal{C}<\mathcal{T}$, rising access $\mathcal{A}$, and a preference for similar types.[11] That is $\mathcal{S}_{ij}$ and μ_A fall, while $E[\mathcal{D}_{ij}]$ rises with agents i and j in different communities.
Proof – Provided in the mathematical appendix.

Corollary 2.1 – Suppose that a group requires at least $\mathcal{M}$ members of the preferred type to achieve critical mass on a focal topic. Then rising access increases the probability that a topic-based or special-interest-group forms.

Proof – This is a straightforward consequence of increasing the sample mean (equation **2.a** in proof of Proposition 2). If $\mathcal{A}(t/\mathcal{N})<\mathcal{M}$ initially, then increasing $\mathcal{A}$ can push the sample mean above critical mass. Subject to agent preferences (cf. Proposition 4), this argument applies also to critical mass defined in percentage terms as defined by $(\mathcal{A}t/\mathcal{N})/\mathcal{C}<\mathcal{M}/\mathcal{C}$. □

Corollary 2.2 – Increasing the number of topics $\mathcal{T}$ in virtual communities reduces integration relative to geographic communities given bounded rationality $\mathcal{C}<\mathcal{T}$ and a preference for similar types.
Proof – Limit behavior of the similarity profile $\mathcal{S}_{ij}$ provides the essential intuition. Taking $\mathcal{T}\to\infty$ (in equation **2.e**) and holding $\mathcal{C}$ constant, causes $\mathrm{Cos}(\Theta_{ij})\to 0$ because $\mathcal{T}^3/\mathcal{T}^4\to 0$. This implies that increasing the number of topics causes the expected profiles to diverge, establishing corollary 2.2. As an aside, it is interesting to note that taking the limit as $\mathcal{C}\to\infty$ gives $\mathrm{Cos}(\Theta_{ij})\to(\mathcal{T}\text{-}2)/(\mathcal{T}\text{-}1)$ which goes to 1 with $\mathcal{T}>\mathcal{C}$. This implies that infinite capacity gives everyone the same knowledge profile, confirming Proposition 1. □

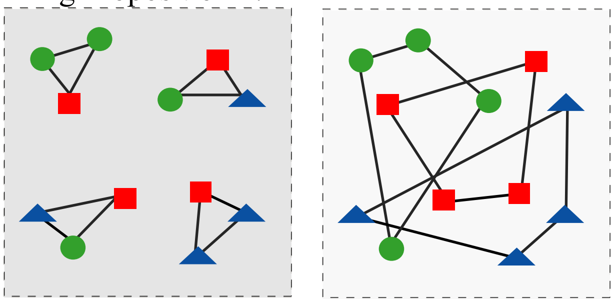


Figure 3.A **Figure 3.B**

As geographic access improves, agents seeking similar types forsake local connections.

[11] The homophily assumption is primarily helpful as illustration and is inessential to our main findings. Later examples will show how preferences consistently favoring a limited number of types confirms Proposition 2. We also endogenize preferences in Proposition 5. If people introduce each other to like contacts whom they know, then narrow contact exposure, even outside one's initial focal area, again gives analogous results.

Figure 3 illustrates geographic connections converting to type connections after global access. In frame 3.A, access is strictly local and bounded by geography, so agents form small communities with fairly uniform knowledge access across types. Frame 3.B shows global access; like types have located one another and formed tightly knit communities of interest. Interconnections across geographic communities have fallen and resources are more concentrated. The integration indices in this example show that for twelve agents and three types, the average $Cos(\Theta_{ij})$ declines from .77 to .27, indicating that profiles diverge; and μ_A falls from .65 to 0, indicating that overlap has fallen among communities. Certain empirical data support the basic proposition. Individuals who spend more time using the Internet have a statistically significant loss of contact with their social environment, and they spend less time with human beings (Nie & Erbring, 2000). They also increase their time spent working at home but do not spend less time with office work. Another study found that local Pittsburgh users decreased their knowledge of Pittsburgh events while their knowledge of national events remained about the same (Kraut, et. al. 2002). Corollary 4, below, will suggest one reason why this is not a necessary consequence of using this technology but of how it is used.

Corollary 2.1 suggests that better communications can strengthen special interest groups, affecting *whether* they form, not just *where* they form. Holding population size constant, rising access helps achieve critical mass. For example, members of hate groups might keep their views private when interacting with their physical neighbors, but they need not do so with like-minded brethren online. More positively, parents of children with leukemia often have too few geographic neighbors to constitute self-help groups, but the likelihood of achieving critical mass rises online. In prior work characterizing the emerging "Global Village," Marshall McLuhan et. al. (1989) recognizes the power of satellite technology to aid "super-regionalisms" and "separatisms." As an historical example, the telephone strengthened affiliation among teenage peer groups (Sproull, 1991). In the field of economics, the number of out-of-state and out-of-country coauthorships in four top journals grew from 4.6% in the 1960s to 27.6% in the 1990s (Gaspar & Glaeser, 1996). Communications scholars have suggested that virtual communities are often more heterogeneous in terms of social attributes (age, income, gender, ethnicity) but more homogenous in terms of attitudes (Hiltz & Wellman, 1997). Similar evidence appears to hold for academia in general:

> "Historically, the strength of an academic department rested with its resident faculty. Now it depends on the extent to which each faculty member is interconnected with other professionals -- worldwide -- pursuing similar interests... We now have electronic research teams and electronic water coolers. This drastically changes – weakens, in my opinion – indigenous workplace relationships and affects workplace cohesiveness."[12]

Contact expansion is also a factor that might influence integration – why would not computer mediated communications (CMC) lift the bounded rationality constraint? The Internet has the attractive property that a person might communicate with a very large group of associates, improving integration. For example, newsgroups and chat rooms have few, if any, physical limitations on participation (Smith, Franham & Drucker, 2002). Broader participation will increase the integration of resources and groups insofar as more people choose to access the same information and insofar as information flows across the boundaries of separate groups. These represent important gains from increased connectivity. This fully parallels Proposition 1.

There are, however, two important qualifications. First, if contact expansion is restricted to members of the same community, then information resources are less fragmented (i.e., μ_I rises) but intergroup fragmentation remains unchanged (i.e., μ_A stays constant). People might also find that the sizes of their groups increase but that the number of group memberships they hold remains relatively constant. By analogy to journal publication, the variety of choices and the number of subscribers might increase, but the number of simultaneous subscriptions held by individuals might not increase and could decrease as a fraction of the total number possible. Due to bounded rationality, the median subscriber base might fall as mass publication gives way to niche publication. This appears to be a strategy of many increasingly focused, or even personalized, net blogs, e-zines, and news feeds. Indeed, online bookstores sell relatively more esoteric books, and relatively fewer mass-market books, than off-line bookstores, reflecting their more sophisticated search and recommendation tools and the broader selection of titles that they provide (Brynjolfsson, Hu and Smith, 2003).

[12]Interview with Edward Mabry, communications researcher at University of Wisconsin, Milwaukee (Leslie, 1995).

Second, participating in a group is not the same as interacting with all of a group's members. The number of subscribers to most newsgroups substantially exceeds the number of people who actually post messages. Joining a new community shortens communications distance (i.e., μ_C rises) only if this establishes new and shorter paths, and it homogenizes information profiles (i.e., μ_I rises) only if members actually communicate. Universal participation is unlikely to imply universal broadcast. Noise and confusion would likely result. Putnam observes that "CMC so lowers the threshold for voicing opinions that, like talk radio, it may not lead to deliberation but to din," further widening the gap between talking and listening (2000, pp 173-74). Expressed differently, as a group gets larger, the fraction of members who post useful communications probably declines after some critical point. In network organizations, for example, sociologists have recognized that with increasing in-group ties, comes decreasing out-group ties due to affinity relations and economizing on time and effort (Baker, 1993) -- an empirical finding that supports the possibility of fragmentation.

At issue is the relative change in communications contacts versus the relative change in capacity. If capacity rises but contacts rise faster, then selectivity rises and people filter more. Although bounded rationality constraints motivate Proposition 2, expanding capabilities to relax this constraint can leave intact the basic results. Graphically, this can be shown in the following manner.

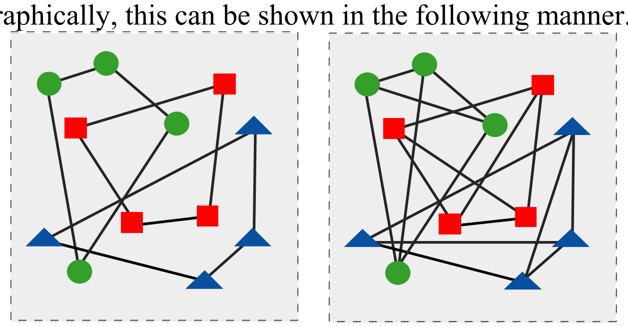


Figure 4.A **Figure 4.B**

Increasing channel capacity does not necessarily increase the measure of integrated affiliations.

In this example, agents use their additional connections to reach additional members of the same community.[13] This leads to the following observation:

[13]This is not to suggest that nothing has changed. Communication paths are (slightly) shorter and information is (slightly) less concentrated implying that μ_C and μ_I have risen. This draws attention to the importance of multiple measures. The Appendix presents an example in which indices even move in opposite directions.

Proposition 3 -- Under global access, relaxing the bounded rationality constraint $\mathcal{C}$ does not integrate agents with different information if they veto communications or if agents choose not to seek topic contacts outside their original areas. Let μ_A' and Θ_{ij}' represent new indices after expanding agents' capabilities. If agents connect to members of the same community, then new indices are unchanged and $\mu_A' = \mu_A$ and $\forall\ i, j$, $\Theta_{ij}' = \Theta_{ij}$ and, in fact, distance $\mathcal{D}_{ij}$ between members of different groups rises.

Proof 3: From Proposition 2, we know that under unrestricted (or "global") access, the knowledge profile of an agent i with preferred contacts, including her own endowment, is $\mathcal{P}_i = [0, 0, \ldots (\mathcal{C}+1)\boldsymbol{\kappa}_t, \ldots 0]$. Let the new number of channels be given by $\mathcal{C}+\Delta$. Then if agents connect to additional knowledge bases of the same type, the new knowledge profile becomes $\mathcal{P}_i = [0, 0, \ldots (\mathcal{C}+1+\Delta)\boldsymbol{\kappa}_t, \ldots 0]$ and similarly for $\mathcal{P}_j = [0, \ldots (\mathcal{C}+1+\Delta)\boldsymbol{\kappa}_s, \ldots 0, 0]$ with $s \neq t$. But then $\text{Cos}(\Theta_{ij}') = \text{Cos}(\Theta_{ij}) = 0$ and since no community has added new members $\mu_A' = \mu_A$ indicating that the respective communities are as just as balkanized as before. Similarly, for unconstrained access, expected agent distance is initially $E[\mathcal{D}_{ij} | t_i \neq t_j] = (\mathcal{C}+1)\boldsymbol{\kappa}_t \sqrt{2}$ while for expanded capacity this becomes $(\mathcal{C}+1+\Delta)\boldsymbol{\kappa}_t \sqrt{2}$. □

New channels would integrate communities if agents used their added capacity to reach outside their original communities. Again, preferences are crucial. In the example above, however, channels have increased by 50% but all agents use their additional resources to communicate with previously unreached members of their existing communities. Resource fragmentation falls (μ_I changes from .19 to .33), but the index of integrated affiliation, μ_A, remains unchanged. Each agent effectively deepens his or her knowledge of a given topic area, rather than broadening it to other topic areas.

VI. The Preferences That Bind

Another feature of Proposition 2 is that agents exhibit strong preferences: they prefer to associate exclusively with agents of like types. Our fourth result relaxes this condition (e.g. some agents might prefer *intentional* randomness) and shows that even weak preferences can lead to similar results. In fact, unless agents are indifferent to their connections or seek *greater* diversity than is locally available, a population with global access will generally increase on measures of balkanization. Let access be unrestricted so that only preferences matter. If t is the prevalence of a given type in a population of $\mathcal{N}$, then define a 'narrower' preference as agent i seeking a higher concentration than $(t/\mathcal{N})$ for some element of $\mathcal{P}_i$. For $\mathcal{C}$ samples, the preferred mean exceeds $\mathcal{C}(t/\mathcal{N})$. We formalize this below.

Proposition 4 – Narrower preferences reduce integration. If an agent prefers narrower associations than an average sample from the population $\mathcal{C}(t/\mathcal{N})$, then balkanization rises.

Stronger preferences lead to greater balkanization with similarity S_{ij} falling and distance $E[D_{ij}| t_i \neq t_j]$ rising.
Proof – Provided in the mathematical appendix.

Preferences, more than technology, drive the main results. That is, technologies create options but preferences enact outcomes. Figure 5 plots statistical averages for 200 societies of 30 agents with three channels and five types. Assuming agents with extra capacity will communicate outside their preferred set, rising access initially drives down balkanization regardless of preferences. As access continues to rise, however, a preference for diversity leads to the greatest integration while a preference for focus leads to the greatest balkanization.

If individuals can choose their content, contacts, and connections, then emphasizing preferred communities can balkanize interactions. Although Figure 5 shows only one type of commitment, this effect does not depend on a preference for a single type. If, for example, a person chooses to interact with a dozen communities when serendipitous geographic interaction would have led to several dozen, then the breadth of exposure to novel types of information is likely to fall.[14] The key point is that if the distribution of tastes differs from the distribution of types in the local population, then technologically-lowered search costs allows people to shift their interactions towards indulging their tastes, possibly lowering levels of integration. With bounded rationality, any preference for membership in virtual communities that is more concentrated than representative geographic samples will reduce affiliations with less favored communities -- the stronger the preferences, the greater the fragmentation. Proposition 4 has powerful implications because it suggests that geography only needs to be more heterogeneous than tastes in order for the lifting of geographic constraints to result in more specialized interactions.

[14]Nor does the result depend on homophilia or a preference for *similar* types. If preferences are such that A will only speak to B who will only speak to C who will only speak to A, then communications deadlock. Likewise, mutual preferences for contact between pairs (A, B) and (C, D) can divide the community into two blocks. In both cases, a preference for narrow types can balkanize interactions.

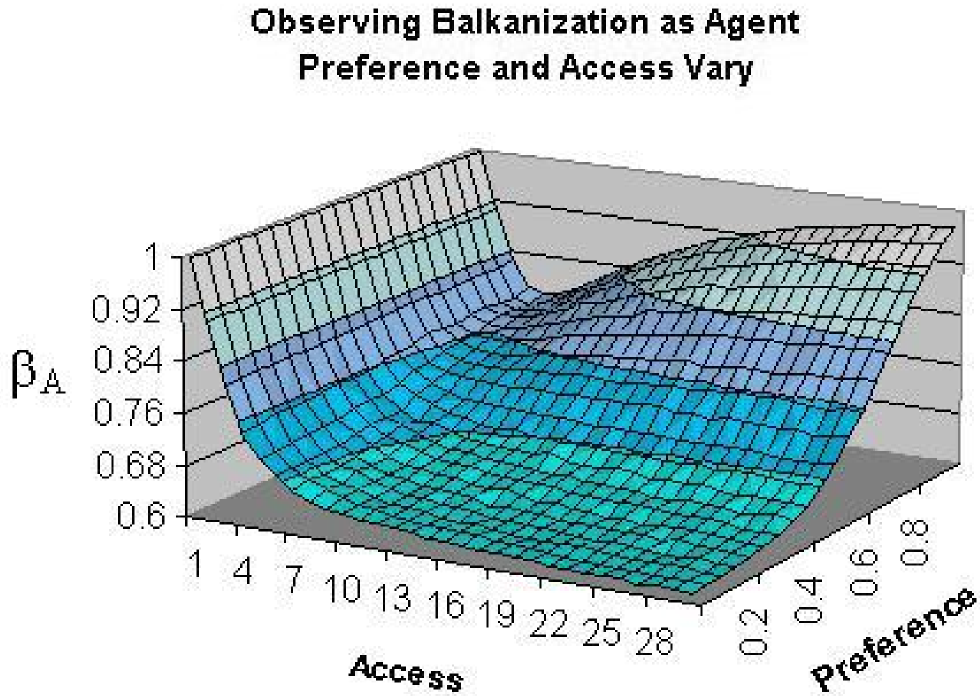


Figure 5 – Under global access, balkanization reaches a maximum for agents with narrow preferences but reaches a minimum for agents with diverse preferences.

As Figure 5 illustrates, specialization and fragmentation are not inevitable. If tastes favor diversity, then the action of preferences reduces specialization and increases integration. In other words, if geography is less heterogeneous than tastes, greater connectivity implies moving from back to front in Figure 5. This gives rise to the following observation:

Corollary 4 -- If agents exhibit a taste for diversity or randomness, which exceeds that available locally, then greater access reduces balkanization.

This corollary simply restates the previous conclusion with tastes running in the opposite direction. The Internet can, in fact, lend itself to experimentation, as when people enter into multi user-domains (MUDs) and alter their virtual personalities and behaviors (Turkle, 1995). Opposites can also attract (Kaufer & Carley, 1993). A preference for how CMC is used also parallels the finding that introverts decrease their community involvement and increase in loneliness while extraverts increase their involvement and decrease in loneliness (Kraut, et. al. 2002). The proposed model can illustrate movement both towards and away from integration. Which effect dominates depends on the distribution of preferences. Narrower preferences lead to specialization while broader preferences lead to integration, as connectivity increases.

VII. Endogenous Preferences: Hyper-specialization?

Relaxing the model further, preferences need not remain static; the desire for affiliation with a particular group can increase or decrease with increased access to technologies like the Internet. To explore this possibility, we introduce a simple model of non-stationary preferences. Let the target level of contact for a given knowledge type be $k_t^* = \sigma(k_t - \bar{k}) + \bar{k}$, where $\bar{k}$ is an agent's average across all knowledge types and k_t is knowledge level for a given type. Sigma $\sigma \in [0,1]$ represents the salience of similar knowledge. With σ=0, agents exhibit a taste for diversity and spread their target capacity uniformly across types. With σ>0, agents prefer associations that play to their strengths and avoid those that play to their weaknesses. That is, they prefer to deepen their knowledge of topics above their average. For instance, an astrophysicist might prefer to read a new article about quasars rather than one about Icelandic syntax, while a linguist might reverse this preference ordering. Below, we model $\sigma \in [0,1]$ although the opposite is easily modeled as $\sigma \in [-1,0]$, representing satiation – agents avoid their areas of expertise.

In a low-tech world with $\mathcal{A} << \mathcal{C}$, agents would like to deepen their knowledge levels but cannot. Preferences are stationary, always positive, and bounded by reach. Interestingly, increasing technological access $\mathcal{A}$ causes target levels to rise across *all* types, even for total focus with a complete salience of similar interest $\sigma = 1$. At first, a Renaissance-like interest in *all* kinds of information appears. These targets rise until the bounded rationality constraint binds. Then, only those targets in which an agent is relatively expert deepen. All other targets fall and, in fact, can fall to zero. Rising access can therefore cause a preference for specialization even after fostering an initial interest in general knowledge. This nicely mirrors Kuhn's (1970) concerns about growing hyperspecialization in academia. We formalize this below.

Proposition 5 – As access $\mathcal{A}$ rises, the expected target for *all* knowledge contact initially rises given $\mathcal{A}<\mathcal{C}$. As access rises past rationality constraint $\mathcal{C}$, expected target contact for similar types continues to rise while the target contact for dissimilar types reverses and falls. As access rises above $\mathcal{C}(\mathcal{N}/t)$, further increases in access have no effect.

Proof – From the fully, partially, and un-constrained equations *2.d*, *2.c* and *2.b*, rising access gives dissimilar types three expected connection strengths $k_d \in \{\mathcal{A}(t/\mathcal{N}),\ C - \mathcal{A}(t/\mathcal{N}),\ 0\}$, and similar types connection strengths $k_s \in \{\mathcal{A}(t/\mathcal{N})+1,\ \mathcal{A}(t/\mathcal{N})+1,\ C+1\}$ respectively. In each case, $\bar{k}$ is given by $(1/\mathcal{T})[(\mathcal{T}-1)k_d + k_s]$ so that algebraic simplification and partial differentiation of $k_i^* = \sigma(k_i - \bar{k}) + \bar{k}$ for $i \in \{s, d\}$ give:

Range	k_d^*	k_s^*	$\frac{\partial k_d^*}{\partial A}$	$\frac{\partial k_s^*}{\partial A}$
$\mathcal{A}<C$	$A\left(\frac{t}{N}\right)+\frac{1-\sigma}{T}$	$A\left(\frac{t}{N}\right)+\sigma+\frac{1-\sigma}{T}$	$\left(\frac{t}{N}\right)$	$\left(\frac{t}{N}\right)$
$\mathcal{A}(t/\mathcal{N})< C < \mathcal{A}$	$\sigma\left(\frac{C-A\left(\frac{t}{N}\right)}{T-1}\right)+\left(1-\sigma\right)\left(\frac{C+1}{T}\right)$	$\sigma\left(A\left(\frac{t}{N}\right)+1\right)+\left(1-\sigma\right)\left(\frac{C+1}{T}\right)$	$\left(\frac{\sigma}{T-1}\right)\left(\frac{-t}{N}\right)$	$\left(\frac{t}{N}\right)$
$C<\mathcal{A}(t/\mathcal{N})$	$\left(1-\sigma\right)\left(\frac{C+1}{T}\right)$	$\sigma(C+1)+(1-\sigma)\left(\frac{C+1}{T}\right)$	0	0

Note the table provides expected connection strengths; for knowledge levels, multiply all terms by κ as in equations *2.d*, *2.c* and *2.b*. Increasing access reverses the sign of the first derivative $\partial k_i^*/\partial \mathcal{A}$ from positive to negative on dissimilar types only. Further, $\partial k_i^*/\partial \mathcal{A} = 0$ once the expected number of contacts of the preferred type $\mathcal{A}(t/\mathcal{N})$ rises above capacity. □

Boundaries on the salience of similarity σ show how targets move. If $\sigma = 1$, the expected equilibrium targets yield exactly equations *2.d*, *2.c* and *2.b*. On the other hand, no preference for similar contact $\sigma = 0$ gives uniform values across all types. More formally:

Corollary 5 – With global access $\mathcal{A}(t/\mathcal{N}) > C$ and complete salience of similarity $\sigma = 1$, contact exposure causes targets to specialize. Target interests converge to a topic singularity with $k_d^* = 0$ for $\mathcal{T}$-1 types and $k_s^* = \kappa(C+1)$ for one. In contrast, with rising access and no salience of similarity $\sigma = 0$, contact exposure causes targets to generalize. Targets diverge to uniform interest across topics with $k_d^* = k_s^* = \kappa[\mathcal{A}(t/\mathcal{N}) + 1/\mathcal{T}]$ for $\mathcal{A}<C$ and $\kappa\,[(C+1)/\mathcal{T}]$ for $\mathcal{A}(t/\mathcal{N}) < C$ and $\kappa\,[(C+1)/\mathcal{T}]$ for $C < \mathcal{A}(t/\mathcal{N})$.

Proof – Substitute $\sigma = 1$ or 0 for k^* in the table above. □

Interpreting the last row of the table, when access falls below capacity, an agent's target for each type of contact is just the expected $\mathcal{A}(t/\mathcal{N})$ value plus a fraction of his own endowment (+1). When access no longer binds, the target becomes his average capacity after accounting for his own endowment.

The proofs specify only expected outcomes. In practice, communities can exhibit bunched distributions of types, so that initial conditions reflect highly non-uniform exposure. Endogenous preferences with an overabundance of multiple types become similarly concentrated on just those types with the proviso that each target is lower than that of an agent focusing on only one type k_t^*. Thus an agent with no endowment but high environmental exposure could become keenly interested in that topic, which, in effect, is the point. Agents need not concentrate only on that which they know initially but, in the context of bounded rationality and a modest degree of focus, they will concentrate on that to which they have been exposed.

IT initially aids search and filtration, but agents may subsequently acquire new tastes or sharpen their preferences. The growth in access in Figures 6.A and B illustrate the condition of positive feedback: an affinity for a particular topic leads an agent to seek either more information from or more solidarity with the community focused on that topic. The target for similar knowledge grows with access. Communications research suggests that "what you know depends on whom you know and who you know depends on whom you meet" (Sproull, 1991 p.11). This can make knowledge profiles path dependent, and it also opens the door to information feedback. Historically, positive feedback is dampened by geography and unfiltered interaction. Communication technology, recommender systems, search engines, and message filters, however, support positive feedback. With "perfect" filtering, σ=1 and the positive feedback target function above, the stopping point is a singularity with all focus on one topic.

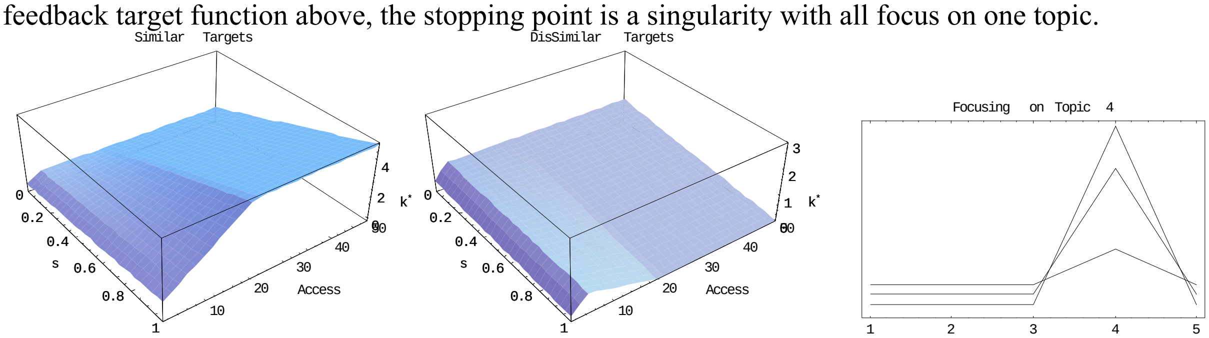


Figure 6A, 6B – Three shaded regions show similar and dissimilar preference changes as access rises through highly constrained ($\mathcal{A}<\mathcal{C}$), partly constrained ($\mathcal{A}(t/\mathcal{N})< \mathcal{C} < \mathcal{A}$), and unconstrained regions ($\mathcal{C}<\mathcal{A}(t/\mathcal{N})$). Figure 6.C shows these same three regions for one focal and four non-focal topics.

The shifts might represent a political dabbler becoming focused on a special interest group or an oenophile graduating from a general interest in wines to a preference for fine burgundies only. Empirical findings appear to support these observations. In illustrating a theory of communication ecology as mutually defining agents, context, and transmissions, Kaufer and Carley (1993) cite several studies in which contact and shared information exhibited a reciprocal relationship. A study of writing students found a positive correlation between their patterns of interaction and emerging consensus. Another study found that employees' similarity, shared information, and proximity predicted their social interactions. Feedback between interaction and shared information also appears to explain differences in observed cohesiveness of certain religious groups (Kaufer & Carley, 1993). Thomas Friedman observed that tribalism and xenophobia coupled with modern media had fostered radicalism:

> Worse, just when you might have thought you were all alone with your extreme views, the Internet puts you together with a community of people from around the world who hate all the things and people you do. And you can scrap the BBC and just get your news from those Web sites that reinforce your own stereotypes. [Friedman, 2002]

Recent technology provides a far higher level of control over interaction and contact filtering. Control exists at the message level that previously extended only to the interpersonal level. Moreover, filtering and screening may take place on behalf of individuals with or without their awareness and consent. News organizations deliver customized news via web browsers, cable services recommend shows based on past viewing habits, and a patent has been issued for software that customizes personalized advertisements over cable channels.[15] Advertising and news stories can target the level of word choice both to spark interest and to penetrate filters designed to screen unwanted contact. For instance, sales tools like collaborative filters typically work by recommending additional products and services which are similar to those purchased or considered in the past. They seem especially effective for “taste” goods like music, reading, and video. Improvements in IT might therefore lead to more focused interaction through the action of choices people make for themselves and the action of choices others make for them. Moreover, if preferences for new information are a function of past information or contacts, then an agent's preferences and connections can become path dependent. Small changes early in community evolution can radically affect their ultimate character.

Many communities offer examples of increasingly narrow focus and specialization. Numerous (most?) academic disciplines, in fact, have progressed to the point where the specialized vocabulary that facilitates interactions within the community hinders interaction across communities. Indeed, Kuhn observed that a widening gulf "separates the professional scientist from his colleagues in other fields" (Kuhn, 1970, p. 21). Specialists in branches of mathematics other than algebraic geometry, for example,

---

[15]US Patent 5,515,098.

have difficulty following the proof of Fermat's last theorem. Differentiation and inbreeding among communities has progressed to the point where expertise can mean “knowing more and more about less and less.” As IT improves filtering, tailoring, segmenting, and searching, the more global network can become the less local village.

VIII. From Specialization to Stratification

The principles that govern integration *across* types also hold *within* types if sufficient differentiation exists to distinguish one classification from another of the same type. Quality can represent one such axis of differentiation. Given quality differences[16] --analogous to type differences -- the same fragmentation can occur. Associations can form among high, middle, and low quality tiers, inducing stratification. As a twist on specialization, this leads to our final proposition.

Proposition 6 -- Quality differentiation in virtual communities can lead to stratification.
Proof -- For specialization due to preferences, the proofs are identical to the proofs of Propositions 2 and 4 with $\mathcal{T}$ interpreted as grades in quality. For specialization due to veto power on the part of a destination community, an alternate proof is available from the authors. □

A graphical representation is identical to Figure 3, with shape interpreted as quality. This interpretation, however, emphasizes one novel dynamic implicit in earlier discussions. Agents at a source might wish to affiliate with agents at a destination, but if agents at the destination have already committed their channels, the destination community is closed. Veto power at a destination can balkanize communities despite preferences for diversity at a source. The publisher of an electronic newsletter, for example, argues

> I hate to sound undemocratic, but if you're going to have valuable discussion, you have to limit it to people with valuable knowledge. The beginners can have their beginner's groups. (Chao, 1995).

New information technologies can therefore exclude as well as enfranchise, as with encryption, intranets, private mailing lists, and firewalls. With respect to quality, agents might wish to connect to others agents' higher-grade resources but find no connections available. Agents in the top tier of an information pyramid might therefore benefit disproportionately from global access. Blau (1977) observes that contact between people of unequal status in one dimension is usually due to a status reversal in

[16] This differs somewhat from Blau (1977) who treats vertical differentiation as “inequality” and horizontal differentiation as “heterogeneity.” Many similar properties emerge, although status distinctions typically apply uniquely to inequality.

another dimension. With rising access, an initial advantage from owning a quality resource grows because it gains an agent possible entry into an otherwise closed community. Importantly, however, closure occurs not from any bias against one group but from a preference for another. There is an implicit competition between choices for interaction only because a limitation on choices means that a decision to include one option results in the exclusion of another.

In fact, this observation is a virtual analog of that provided by Schelling's (1977) model of community separation. If everyone chooses neighborhoods where at least 51% of their neighbors are the same type, then neighborhoods can become segregated and homogenized even though no one objects to a 49% prevalence of a different type. Each simply prefers not to be in the minority. It is straightforward to carry this insight into virtual space where geographic separation in 2-dimensions can manifest itself in N-dimensions. Separation can rise as the dimensionality rises.

A combination of stratification and specialization are depicted in Figure 8. Members of communities A and B might be geographically close (8.A) but have knowledge of diverse topics A' and B' (8.B). Global networks enable new communities A' and B' which are geographically diverse but specialize with respect to topic area (8.B). Finally, 8.C shows how adding other axes of differentiation creates an opportunity for further fragmentation, including stratification. Such axes can include, for example, subspecialization within an existing discipline.

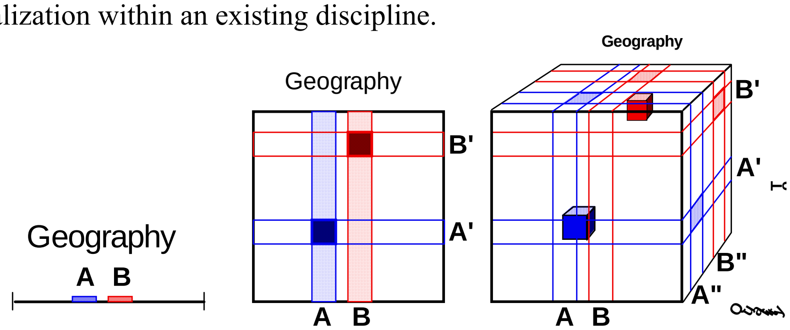


Figure 8.A **Figure 8.B** **Figure 8.C**

Geography's influence on near neighbors diminishes as other dimensions such as topic and quality come to the fore. Fragmentation can increase with each new dimension, progressing from geographic separation (8.A), to topic specialization (8.B), to quality stratification (8.C).

Focusing interaction can homogenize *intra*-action but at the cost of separating groups. Figure 8.C, in particular, shows how one model of community interactions could place near geographic neighbors in different topic camps. If IT shrinks distance, spending time "abroad" can imply neighbors might become strangers and measures of knowledge overlap might fall with rising fragmentation of communities.

IX. Conclusions and Implications: Should We Care About Balkanization?

In this paper, we have defined measures of integration, we have developed a model of possible affiliation based on individual preferences, and we have used these tools to explore possible theoretical implications of changes wrought by IT. In particular, these changes affect our capacity to select, search, screen, and connect. As these abilities influence knowledge profiles and community membership, they also influence the diversity and integration of the communities we voluntarily form. Our findings are suggestive of possible future scenarios and of paths dependent on the interactions we choose for ourselves. In this context, conditions set forth in our various propositions help to guide the choices we might wish to make in light of such factors as diversity, equality, and efficiency. Furthermore, the metrics we define can be used both to advance theoretical analyses and as the basis for empirical assessments of the effects of electronic connectivity.

Under certain conditions, specialization and fragmentation can be economically efficient and stable in the sense that no individual can be made better off by changing their personal affiliations from those under focused interaction. It might even be possible for this to be welfare-maximizing in the sense that specialized production may produce the greatest output. Interconnected "collaboratories" can allow scientists to share data and access remote instruments (Wulf, 1993). Oceanographers who converse through communications technology are more productive -- they author more papers, earn greater peer recognition, and know more colleagues (Hesse, et. al., 1993). Business function diversity can increase measures of team innovation at the expense of implementation efficiency (Ancona & Caldwell, 1992). And, economists sometimes assert that matching peers is efficient (Kremer, 1993; Roth, 1984), as the benefits of specialization date back to the work of Adam Smith (1776).

Voluntary fragmentation, however, might also prove destructive to the overall welfare of society due to over-specialization, and it has been widely criticized by lawyers, scientists, and social commentators. Posner (2001) criticizes modern public intellectuals for stepping from specialized academic disciplines into public arenas where they make proposals that are silly -- or worse, counterproductive -- from a deeper, more well-balanced, and interdisciplinary perspective. Wilson (1998) uses the term "consilience" to describe the unification of knowledge across physical and social sciences, under which alternate explanations are connected to and consistent with one another. He argues that "professional atomization" destroys this unification, leading to weak understanding and harmful relativism. Burke (1996) connects seemingly unrelated ideas and events that subsequently give rise to major scientific breakthroughs. Common to each of these observations is the idea that increased intellectual depth can incur serious costs in terms of breadth.

In fact, Watson and Crick combined skills from zoology and x-ray diffraction to determine the structure of DNA (Moffat, 1993). Thomas Kuhn developed his ideas on scientific paradigm shifts while working at the nexus of history and physics; yet it would be extremely difficult to look for common principles across paradigms by examining a single paradigm. Once Black and Scholes recognized their formula for options pricing as a physics equation for heat transfer (Black & Scholes, 1973, p. 644) they could look for established parallels. Similarly, the Alvarez theory that an asteroid caused the extinction of the dinosaurs emerged from a fortuitous combination of father and son skills in astrophysics and geology (Alvarez, 1980). "Some of the greatest achievements in science come from work at the boundaries of disciplines."[17] The difficulty is that it is not always clear beforehand which groups need to share information.

Even in those cases where the connections are obvious, the necessary interactions may fail to occur. If the returns to individuals from more specialized interaction do not align with the returns to societies, then persons acting out of pure self-interest will not internalize the spillover effects -- the externalities, say, from nonrival information transfers -- that would otherwise benefit society. Voluntary fragmentation might then produce direct economic costs. A reduction in face-to-face interactions between neighbors, for example, presages increased crime rates affecting entire neighborhoods (Putnam, 2000). Similarly, the intellectual benefits of "cross-pollination" between academic researchers and business practitioners might spill over into government policy. In the presence of such externalities, groups that do

[17](Angier, 1985, citing Robert Hazen)

not internalize the benefits they create for others will interact too infrequently. In such situations, reduced specialization would improve social welfare.[18]

Independent of the potential economic costs or benefits of integration, members of a society may wish to increase integration simply to maintain a degree of social cohesiveness. With the customized access and search capabilities of IT, individuals can focus their attention on career interests, music and entertainment that already match their defined profiles, and they can arrange to read only news and analysis that align with their preferences. Individuals empowered to screen out material that does not conform to their existing preferences may form virtual cliques, insulate themselves from opposing points of view, and reinforce their biases. Authors of collaborative filtering technology have long recognized its ability to both foster tribalism as well as a global village (Resnick, 1994).

Indulging these preferences can have the perverse effect of intensifying and hardening pre-existing biases (Sunstein, 2002). Thus people who oppose free trade are likely, after talking to one another, to oppose it more fiercely; people who fear gun control appear, after discussion, more likely to take action; and juries that want to send a message seem, after deliberation, to set higher damage awards. The reasons include information cascades and oversampled arguments. In one, an accumulating, and unchallenged, body of evidence leads members to adopt group views in lieu of their own. In the other, members of a limited argument pool are unwilling or unable to construct persuasive counter-arguments that would lead to more balanced views. The effect is not merely a tendency for members to conform to the group average but a radicalization in which this average moves toward extremes (Sunstein, 2002). Increasing the number of information sources available may worsen this effect, as may increasing the attention paid to these information sources. For instance, according to a study by Kull et al. (2003), of 8634 respondents, those with strong political opinions were more likely to have misperceptions about key facts regarding the Iraq war, and this effect was exacerbated among people who paid more attention to the news. Republicans and Democrats selectively preferred different news sources and these sources tended

[18]This assumes that there is a "missing market" for externalities, since a third community could presumably subsidize any interaction that provided measurable benefits. In the case of "cross-pollination," one information paradox is that a potential buyer cannot accurately assess the value of shared information without inspecting it, but having inspected it, the buyer cannot in good faith return it to the seller and claim also to know nothing of what he has seen (Arrow, 1962). It therefore seems probable that both externalities and missing markets exist.

to provide information that reinforced pre-existing beliefs. As the authors noted “Higher levels of exposure to news compound[ed] the effect of political positions on the frequency of misperceptions” (p. 19)

Internet users can seek out interactions with like-minded individuals who have similar values and thus become less likely to trust important decisions to people whose values differ from their own. This voluntary balkanization and the loss of shared experiences and values may be harmful to the structure of democratic societies as well as decentralized organizations. In the U.S., Putnam’s (2000) influential book has documented a substantial decline in social and civic engagement over the past thirty-five years. He reports that citizens vote less, attend church less, socialize with neighbors less, and volunteer less. In contrast, he argues that civic participation correlates with lower crime rates, greater political responsiveness, and better schools. Econometric evidence suggests that inequality and ethnic fragmentation significantly reduce social activity (Alessina & Ferrara, 2000). The gap between information rich and information poor can also widen with virtual communities (Van Alstyne & Brynjolfsson, 1995). These observations augment survey findings that Internet use reduces time with immediate family and local friends and decreases local knowledge (Nie & Erbring, 2000; Kraut, et. al. 2002). Our observation is that preferences represent one of the levers affecting these outcomes.

If diversity of interaction or equality of resources represent goals we would ascribe to our social planners, we need to consider what level of integration we deem most suitable for balancing our private interests as individuals and our shared interests as members of a community. Fragmentation in one or more dimensions of our interactions may or may not be desirable, but once achieved, it can be difficult to reverse. In any event, at this relatively early stage of developing information infrastructure, no single scenario is inevitable. We can, and should, explicitly consider what we value as we shape the nature of our networks and infrastructure — with no illusions that a greater sense of community will inexorably result.

XII. Bibliography

Abler, R., Adams, J., & Gould, P. (1971). Spatial Organization: The Geographers View of the World. Englewood Cliffs, NJ: Prentice Hall.

Alesina, Alberto and Eliana La Ferrara. (2000). "Participation in Heterogeneous Communities." The Quarterly Journal of Economics CXV (3): 847-904.

Allen, T. (1984). Managing the Flow of Technology. Cambridge, Mass.: MIT Press.

Ancona, D.G. & Caldwell D.F. (1992). "Demography and Design: Predictors of New Product Team Performance" Organization Science, 3: 321-341.

Alvarez, L. W., Alvarez, W., Asaro, F., & Michel, H. V. (1980). "Extraterrestrial Cause for the Cretaceous-Tertiary Extinction -- Experimental Results and Theoretical Interpretation." Science, 208(4448): 1095-1108.

Angier, N. (1985, May 6). "Did Comets Kill the Dinosaurs?" Time.

Arrow, K. (1962). "Economic Welfare and the Allocation of Resources for Invention." In R. Nelson (Ed.), The Rate and Direction of Inventive Activity, (pp. 609-625). New Jersey: Princeton University Press.

Banks, David L. and Kathleen M. Carley (1996). "Models for Network Evolution." Journal of Mathematical Sociology. 21(1-2): 173-196.

Baker, W. E. (1993). The Network Organization in Theory and Practice. In N. Nohria & R. G. Eccles (Eds.), Networks and Organizations, (pp. 397-429). Boston, MA: Harvard Business School Press.

Black, F., & Scholes, M. (1973). The Pricing of Options and Corporate Liabilities. Journal of Political Economy, 81(3): 637-654.

Blau, Peter (1977). Inequality and Heterogeneity. New York: Free Press.

Borko, H., & Menou, M. J. (1983). Index of Information Utilization Potential: The Final Report of Phase II of the I.U.P. Pilot Project . Paris: UNESCO.

Brynjolfsson, E., M. D. Smith, and Y. Hu (2003) "Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers" Management Science, Vol. 49, No. 11

Burke, J. (1997) The Pinball Effect: How Renaissance Water Gardens Made the Carburetor Possible. Back Bay Books.

Burt, R. (1993). "The Social Structure of Competition." In N. Nohria & R. G. Eccles (Eds.), Networks and Organizations, (pp. 57-91). Boston, MA: Harvard Business School Press.

Byrne, D. (1971) The Attraction Paradigm, New York: Academic Press.

Cairncross, F. (1997). The Death of Distance: How the Communications Revolution will Change our Lives. Boston, MA: Harvard Business School Press.

Carley, K. (1990). "Group Stability: a Socio-Cognitive Approach." pp. 1-44 in Lawler E., Carley, K. (1991). "A Theory of Group Stability," American Sociological Review 56(3): 331-354.

Carley, K. (1995). "Communication Technologies and Their Effect on Cultural Homogeneity, Consensus, and the Diffusion of New Ideas." Sociological Perspectives, 38 (4): 547-571 Markovsky B., Ridgeway C. & Walker H. (Eds.) Advances in Group Processes: Theory and Research. Vol VII. Greenwhich, CN: JAI Press.

Chao, J. (1995, June 7). Internet Pioneers Abandon World They Created. Wall Street Journal: B8.

Coffman, K.G. and Andrew Odlyzko. (2002). The Size and Growth Rate of the Internet. http://fisrtmonday.org/issues/issue3_10/coffman/index.html

Dodds, P. S., R. Muhamad and D. J. Watts."An Experimental Study of Search in Global Social Networks," by Science, Aug 8 2003.

Donath, J. (1995, Nov. 5-9). Visual Who: Animating the Affinities and Activities of an Electronic Community. Paper presented at the ACM Multimedia, San Francisco, CA.

Fink, Edward L. (1996). Dynamic Social Impact Theory and the Study of Human Communication. Journal of Communication 46 (4): 4-12.

Freeman, L. C. (1979). Centrality in Social Networks: Conceptual Clarification. Social Networks, 1: 215-239.

Friedman, Thomas. L. (2002) "Global Village Idiocy" New York Times, May 12, http://www.nytimes.com/2002/05/12/opinion/12FRIE.html

Gaspar, J., & Glaeser, E. L. (1996). Information Technology and the Future of Cities : NBER.

Granovetter, M. (1973). The Strength of Weak Ties. American Journal of Sociology, 78(6), 1360-1380.

Habermas, J. (1987). The Theory of Communicative Action (McCarthy, Thomas, Trans.). (Vol. I). Boston: Beacon Press Books.
Hesse, B., Sproull, L., Kiesler, S., & Walsh, J. (1993). "Returns to Science: Computer Networks in Oceanography." Communications of the ACM, 36(8): 90-101.
Hiltz, S. R., & Wellman, B. (1997). "Asynchronous Learning Networks as a Virtual Classroom." Communications of the ACM, 40(9): 44-49.
Kaufer, D. S., & Carley, K. M. (1993). Communication at a Distance: The Influence of Print on Sociocultural Organization and Change. Hillsdale, NJ: Lawrence Erlbaum Associates.
Kraut, Robert et al. (2002). Internet Paradox Revisited. Journal of Social Issues 58 (1): 49-74.
Kremer, M. (1993). The O-Ring Theory of Economic Development. Quarterly Journal of Economics, 108(3): 551-575.
Kuhn, T. (1970). The Structure of Scientific Revolutions. (2nd ed.). Chicago: University of Chicago Press.
Kull, S., Ramsay, C., Subias, S., Lewis, E., Warf, P. (2003) Misperceptions, the Media and the Iraq War PIPA/Knowledge Networks Poll, University of Maryland.
Latane, Bibb. (1996). "Dynamic Social Impact: The Creation of Culture by Communication." Journal of Communication 46 (4): 13-11.
Latane, B. and MJ Bourgeois (1996) "Experimental evidence for dynamic social impact: the emergence of subcultures in electronic groups" Journal of communication 46(4): 35-47
Leslie, J. (1995). Mail Bonding. Wired, 2: 42-48.
Malone, T. W., & Smith, S. A. (1988). "Modeling the Performance of Organizational Structures." Operations Research, 36(3): 421-436.
Manning, C. & Schutz, H. (2000). Foundations of Statistical Natural Language Processing. Cambridge, MA: MIT Press.
McLuhan, M., & Powers, B. R. (1989). The Global Village -- Transformations in World Life and Media in the 21st Century. New York: Oxford University Press.
Moffat, A. S. (1993). "New Meetings Tackle the Knowledge Conundrum." Science, 259: 1253-55.
National Technology and Information Administration (1993). The National Information Infrastructure: Agenda for Action : Commerce Department.
Nie, N. & Erbring, L. (2000). "Internet and Society – A Preliminary Report" Stanford Working Papers http://www.stanford.edu/group/siqss.
O'Reilly, C., Caldwell, D. and W. Barnett, (1989). " Work Group Demography, Social Integration and Turnover," Administrative Science Quarterly, 34 21-37.
Orlikowski, W. J. (1992). "Learning from Notes: Organizational Issues in Groupware Implementation," in Conference on Computer Supported Cooperative Work. J. Turner and R. Kraut. Toronto, Association for Computing Machinery: 362-369.
Posner, R. (2003). Public Intellectuals: A Study of Decline. Harvard University Press.
Putnam, R (2000). Bowling Alone. New York: Touchstone Press.
Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Paper presented at the Computer Supported Cooperative Work.
Roth, A. E. (1984). "The Evolution of the labor Market for Medical Interns and Residents: A Case Study in Game Theory." Journal of Political Economy, 92(6): 991-1016.
Schelling, T. (1978) *Micromotives and Macrobehavior*, Norton.
Sanil, A., Banks, D., & Carley, K. (1995). "Models for Evolving Fixed Node Networks: Model Fitting and Model Testing. Social Networks 17: 65-81
Simon, H. A. (1957). Models of Man, Social and Rational. New York: John Wiley & Sons.
Smith, A. (1776) The Wealth of Nations. Penguin Classics.
Smith, M. A. (1999). Invisible Crowds in Cyberspace: Mapping the Social Structure of the Usenet. Communities in Cyberspace: Perspectives on New Forms of Social Organization. London, Routledge Press.

Smith, M.A., Farnham, S.D. & Drucker, S.M. (2002). The Social Life of Small Graphical Chat Spaces, Working Paper, Microsoft Research, Redmond, WA.
Sproull, L., & Kiesler, S. (1991). Connections -- New Ways of Working in the Networked Organization. Cambridge: MIT Press.
Sunstein, C (2002) "The Law of Group Polarization," Journal of Political Philosophy 10:175.
Teachman, J.D.(1980). "Analysis of Population Diversity: Measures of Qualitative Variation," Sociological Methods and Research, 8: 341-362.
Turkle, S. (1995). Life on the Screen -- Identity in the Age of the Internet. New York: Simon & Schuster.
Valente, T.W. (1995). Network Models of the Diffusion of Innovations. Cresskill, NJ: Hampton Press.
Van Alstyne, M. (1997). The State of Network Organization: A Survey in Three Frameworks. Journal of Organizational Computing, 7(3): forthcoming.
Van Alstyne, M., & Brynjolfsson, E. (1995, December 14-17). Communication Networks and the Rise of an Information Elite -- Do Computers Help the Rich get Richer? Paper presented at the International Conference on Information Systems, Amsterdam.
Van Alstyne, M & Brynjolfsson, E. (1996, November 29). "Could the Internet Balkanize Science?" Science 274(5292): 1479-1480.
von Hippel, E. (1998) "Economics of Product Development by Users: the Impact of 'Sticky' Local Information" Management Science 44(5): 629-644.
Wasserman, S. & Faust, K. (1994). Social Network Analysis: Methods and Applications. Cambridge, UK: Cambridge University Press.
Watts, D. & Strogatz, S. (1998, June). "Collective Dynamics of Small World Networks" Nature 393(4): 440-442.
Wellman, B., & Gulia, M. (1997). Net Surfers Don't Ride Alone: Virtual Communities as Communities. In P. Kollock & M. Smith (Eds.), Communities in Cyberspace, (Vol. (in Press). Berkeley: University of California.
Wellman, Barry and Scot Wortley, (1990). "Different Strokes from Different Folks: Community Ties and Social Support." American Journal of Sociology 96 (Nov.): 558-88.
White, H. C., Boorman, S. A., & Breiger, R. L. (1976). Social Structure from Multiple Networks I -- Blockmodels of Roles and Positions. American Journal of Sociology, 81(4):730-780.
Wilson, E. O. (1999) Consilience – The Unity of Knowledge. Vintage Books.
Wulf, W. A. (1993). The Collaboratory Opportunity. Science, 261(5123: 854-855.
Wyner, G. and T. Malone (1996). Cowboys or Commanders: Does Information Technology Lead to Decentralization? International Conference on Information Systems, Cleveland, OH, ACM.
Zipf, G. (1946). "Some Determinants of the Circulation of Information," American Journal of Psychology59:401-421.

XIII. Mathematical Appendix

Proof of Proposition 2 – We show three results: knowledge profiles for agents *i* and *j* in different communities diverge with rising access, μ_A falls, and $E[\mathcal{D}_{ij} | t_i \neq t_j]$ rises. For geographically restricted access, we show $\mathrm{Cos}(\Theta_{ij}) \geq 0$ initially,[19] implying their knowledge profiles overlap. For geographically unrestricted access, we show $\mathrm{Cos}(\Theta_{ij}) = 0$ implying that profiles are completely orthogonal.
Let t be the prevalence of a given type in a population of $\mathcal{N}$ when access provides $\mathcal{A}$ samples. Then availability of each type is a hypergeometric probability distribution with probability of χ contacts to this type given by:

[19] Only if geographic communities are completely balkanized initially will $\mathrm{Cos}(\Theta_{\iota\varphi}) = 0$. Any level of prior integration during initial conditions implies the inequality is strict.

$$\frac{\binom{t}{x}\binom{N-t}{A-x}}{\binom{N}{A}}$$

2.a

The mean is $E[x] = \mathcal{A}(t/\mathcal{N})$. Impose preferences on $E[x]$ given $\mathcal{C}$.. For geographically unrestricted access, $\mathcal{A}(t/\mathcal{N}) \geq \mathcal{C}$ thus agents can expect to use all channels to contact their preferred type. Given i's own resources, the expected knowledge profile is:

2.b $$P_i = \langle 0,0,\ldots(C+1)\kappa,0,\ldots 0\rangle \text{ where } \mathcal{A}(t/\mathcal{N}) \geq \mathcal{C}$$

and similarly for j with $(\mathcal{C}+1)\kappa$ occurring in a different slot. The dot product of these vectors must be zero so $\mathrm{Cos}(\Theta_{ij}) = 0$ after unrestricted access.[20]

For geographically restricted access, either the mean is below capacity $\mathcal{A}(t/\mathcal{N}) < \mathcal{C}$, or access itself is below capacity, $\mathcal{A} < \mathcal{C}$. In the first case, including an agent's own resources, there will be $\mathcal{A}(t/\mathcal{N})+1$ expected contacts of the preferred type, and $\mathcal{C} - \mathcal{A}(t/\mathcal{N})$ contacts spread over $\mathcal{T}$-1 indifferent types. The expected profile is:

2.c $$P_i = \left\langle \frac{[C-A\frac{t}{N}]}{(T-1)}\kappa, \frac{[C-A\frac{t}{N}]}{(T-1)}\kappa, \ldots [A\frac{t}{N}+1]\kappa, \frac{[C-A\frac{t}{N}]}{(T-1)}\kappa, \ldots \right\rangle \text{ where } \mathcal{A} \geq \mathcal{C} > \mathcal{A}(t/\mathcal{N})$$

In the second case, capacity can accommodate all expected contacts $\mathcal{A}(t/\mathcal{N})$.

2.d $$P_i = \left\langle (A\frac{t}{N})\kappa, (A\frac{t}{N})\kappa, \ldots (A\frac{t}{N}+1)\kappa, (A\frac{t}{N})\kappa, \ldots \right\rangle \text{ where } \mathcal{C} > \mathcal{A}$$

Equation 2.b represents the *un*constrained, 2.c the *partly* constrained, and 2.d the *fully* constrained cases, respectively. Given 2.c, the numerator of $\mathrm{Cos}(\Theta_{ij})$ is given by $\mathcal{T}$-2 indifferent terms and 2 preferred terms. The denominator of $\mathrm{Cos}(\Theta_{ij})$ has $\mathcal{T}$-1 indifferent terms and only 1 preferred term. Simplifying and canceling terms in the ratio of $\mathcal{S}_{ij} = \mathrm{Cos}(\Theta_{ij})$ yields:

2.e $$\frac{\mathrm{P}_i \bullet \mathrm{P}_j}{\|\mathrm{P}_i\|\|\mathrm{P}_j\|} = \frac{(T-2)(C-A\frac{t}{N})^2 + 2(T-1)(C-A\frac{t}{N})(A\frac{t}{N}+1)}{(T-1)(C-A\frac{t}{N})^2 + (T-1)^2(A\frac{t}{N}+1)^2}$$

Note that the only interesting cases require $\mathcal{T} > 2$, and that due to restricted access the expected number of contacts $\mathcal{A}(t/\mathcal{N}) < \mathcal{C}$ so $\mathcal{C}-\mathcal{A}(t/\mathcal{N}) > 0$. Thus all terms are non-negative, establishing the result that $\mathrm{Cos}(\Theta_{ij}) \geq 0$ and knowledge profiles overlap under restricted access. The proof for 2.d is similar and implies even greater overlap under more restricted access.

Expected distances $E[\mathcal{D}_{ij} | t_i \neq t_j]$ in the unconstrained, partly constrained, and fully constrained cases are:

---

[20] This argument ignores integer boundary conditions. It is quite possible that remainder $r = (\mathcal{C}+1)\mathrm{mod}(\mathcal{N}) \neq 0$ so that r agents cannot meet their preferences. Assuming $\mathcal{N}$ is large, however, drives the fraction $r/\mathcal{N}$ of such agents toward 0.

2.f
$$D_{ij} = \begin{cases} \sqrt{0,0,...(C+1)^2\kappa^2,(C+1)^2(-\kappa)^2,0,..} & A\left(\frac{t}{N}\right) \geq C \\ \sqrt{0,0,...\left([A\frac{t}{N}+1]-\frac{[C-A\frac{t}{N}]}{(T-1)}\right)^2\kappa^2,\left(\frac{[C-A\frac{t}{N}]}{(T-1)}-[A\frac{t}{N}+1]\right)^2\kappa^2,0,..} & A \geq C > A\left(\frac{t}{N}\right) \\ \sqrt{0,0,...\kappa^2,(-\kappa)^2,0,..} & C \geq A \end{cases}$$

In order, these simplify to $(C+1)\kappa(\sqrt{2})$, $[(1/(T-1))[TA(t/N)-C]+1]\kappa(\sqrt{2})$, and $\kappa(\sqrt{2})$. Plugging in the boundary conditions shows that distances are falling in reduced access as preferred contacts are increasingly constrained.

For μ_A it is easier to consider the probability that two agents with different interests will join each other's communities. As access rises, this probability falls. The initial probability that *i* does not contact any member of j's community is $1 - [C - A(t/N)]/(T-1)$ and similarly for *j*. Thus, for sufficiently large populations, the probability that *j* and *i* are in different communities is

2.g
$$\left[1-\frac{(C-A\frac{t}{N})}{(T-1)}\right]^2$$

As access A rises, the expected number of preferred contacts $A(t/N)$ eventually exceeds C so that the probability *i* and *j* join different communities goes to 1. The balkanization index μ_A must therefore rise with increasing access, completing the proof. □

Proof of Proposition 4 – We show that allocating more channels to one type causes knowledge profiles to diverge in different communities. If agents are indifferent to their connections, concentration will be $C(t/N)$ for any A and C (if access binds instead of capacity, simply replace C with A below). With little loss in generality, we can simplify analysis by allowing the κ_{it} to equal a constant κ and each of the various types *t* be equiprobable. Since there are a total of T types, the likelihood of drawing a type *t* is $(t/N) = (t/Tt) = (1/T)$. The expected number of contacts by type is thus (C/T). Since agents reach their own knowledge bases with certainty, the expected knowledge profile of an agent *i* is $P_i = \langle (C/T)\kappa, (C/T)\kappa, ... (1+C/T)\kappa, ... (C/T)\kappa \rangle$. For *i* and *j* in different communities, we have $\mathrm{Cos}(\Theta_{ij})$ =

4.a
$$\frac{P_i \bullet P_j}{\|P_i\|\|P_j\|} = \frac{(T-2)\left(\frac{C}{T}\right)^2 k^2 + 2\left(\frac{C}{T}\right)\left(1+\frac{C}{T}\right)k^2}{(T-1)\left(\frac{C}{T}\right)^2 k^2 + \left(1+\frac{C}{T}\right)^2 k^2}$$

With algebraic simplification, the κ's cancel and this expression reduces to:

4.b
$$\frac{2C+C^2}{2C+C^2+T}$$

This measures overlap if agents are equally happy mixing with the population at large. If, on the other hand, agents are not indifferent to their connections but prefer to allocate X of their channels to a specific type other than *j* and *j*'s preference, then the expression for random association becomes

4.c
$$\frac{2(C-X)+(C-X)^2}{2(C-X)+(C-X)^2+T}$$

For all $\mathcal{X}$, $1 \leq \mathcal{X} \leq \mathcal{C}$, this implies that overlap between profiles falls. Note that if agents allocate all channels to preferred types, then $\mathcal{X}=\mathcal{C}$ leading again to complete balkanization. Further, these channels need not be reserved only for *like* types; they need only be narrowly allocated. This establishes the point that stronger agent preferences lead to greater balkanization.

We can actually take these results a step further. Using the results from **2.e**, we can show the unusual result that severely restricting access has the same effect as imposing indifference. For severely restricted access, $\mathcal{A} < \mathcal{C}$ so that $\mathcal{A}$ binds in **2.e**. Substituting $\mathcal{A}$ for $\mathcal{C}$ and using the same simplification to replace all instances of $t/\mathcal{N}$ with $\mathcal{T}$, then **2.e** reduces to:

4.d
$$\frac{2A + A^2}{2A + A^2 + T}$$

which resembles the equation for indifferent connections **4.b** with the caveat that all agents are below their capacity. Thus, more restricted access can induce more diverse interaction. Note also that neither the results of Proposition 2 nor those of 4 depend on a homogeneous population distribution; non-uniform clustering gives similar results. In this example, setting $\mathcal{T}=t/\mathcal{N}$ makes the equations more tractable, but it only needs to be the case that some capacity is used to contact different types under restricted access for balkanization to rise with strong preferences under increased access. It is not necessary that types be uniformly distributed.

To prove that narrower preferences increase expected distance $\mathcal{D}_{ij}$, note that reserving channels creates profile $\mathcal{P}_i = \langle((\mathcal{C}\text{-}\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}, ((\mathcal{C}\text{-}\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}, \ldots (1+\mathcal{X}+(\mathcal{C}\text{-}\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}, \ldots ((\mathcal{C}\text{-}\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}\rangle$ with expected distance $E[\mathcal{D}_{ij} | t_i \neq t_j] = \sqrt{2}\,(1+\mathcal{X}+(\mathcal{C}\text{-}\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}$ which increases in $\mathcal{X}$. □