Global Village or Cyber-Balkans?

Modeling and Measuring the Integration of Electronic Communities

Marshall Van Alstyne
Boston University & MIT
595 Commonwealth Ave. Rm 634
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

XIII. Mathematical Appendix

Proof of Proposition 2 – For access rising past capacity, we show three results: knowledge profiles for agents i and j in different communities diverge, μ_A falls, and $\mathcal{D}(E[\mathcal{P}_i], E[\mathcal{P}_j] \mid t_i \neq t_j)$ rises. For geographically restricted access $0 < \mathcal{A} < \mathcal{C}$, we show $\mathrm{Cos}(\Theta_{ij}) \geq 0$ initially,[1] implying their knowledge profiles overlap. For geographically unrestricted access, we show $\mathrm{Cos}(\Theta_{ij}) = 0$ implying that profiles are completely orthogonal. To avoid exotic boundary cases, let the population $\mathcal{N}$ be large and in addition to $t_i \neq t_j$, there exists at least one other type so $T > 2$.

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 x contacts to this type given by:

2.a

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

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.[2]

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

[2] 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.

For geographically restricted access, we show $\text{Cos}(\Theta_{ij}) > 0$. 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 $\text{Cos}(\Theta_{ij})$ is given by $\mathcal{T}$-2 indifferent terms and 2 preferred terms. The denominator of $\text{Cos}(\Theta_{ij})$ has $\mathcal{T}$-1 indifferent terms and only 1 preferred term. Working with expected values, individual vector elements are constant. Simplifying and canceling terms in the ratio of $\mathcal{S}_{ij} = \text{Cos}(\Theta_{ij})$ yields:

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

All terms are non-negative, establishing the result that $\text{Cos}(\Theta_{ij}) \geq 0$ and knowledge profiles overlap under restricted access. This remains true even if we allow non-constant or non-independent random draws on vector element *x*. The proof for 2.d is similar and implies even greater overlap under more restricted access.

Then, given fixed expected profile values, distances $\mathcal{D}_{ij} = \mathcal{D}(\mathrm{E}[\mathcal{P}_i], \mathrm{E}[\mathcal{P}_j] \mid t_i \neq t_j)$ in the unconstrained, partly constrained, and fully constrained cases are:

2.f

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

In order, these simplify to $(C+1)\boldsymbol{\kappa}(\sqrt{2})$, $[(1/(T-1))[TA(t/N)-C]+1]\boldsymbol{\kappa}(\sqrt{2})$, and $\boldsymbol{\kappa}(\sqrt{2})$. 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. Given that *i* contacts it's own type first, the probability that *i's* first random draw is not of *j's* type is $(T-2)/(T-1)$. The number of such random draws is $[C- A(t/N)]$ and similarly for *j*. Thus, for sufficiently large populations, the probability that *j* and *i* are in different communities is

2.g $$\left(\frac{T-2}{T-1}\right)^{2\left(C-\frac{At}{N}\right)}$$

Rising access A shrinks the exponent on this fraction such that under homophily the probability *i* and *j* join different communities goes to 1.

In fact, intuition for the entire proof is straightforward under a simple connection game. Imagine that connections are formed under local access and agents can change them as access increases. As each new connection becomes available, it will be ignored if both agents are of different types or if the target same-type agent has no free channels. With no connection changes, metrics do not change. In contrast, if a same type agent becomes available who is not saturated, each drops an off-type neighbor in favor of the same type match. This change monotonically increases balkanization.[3] □

Proof of Proposition 4 – We show that targeting more channels to one type causes knowledge profiles to diverge for untargeted communities (profiles merge for targeted communities). If agents are indifferent to their connections, expected concentration is $\mathcal{C}(t/\mathcal{N})$ for any $\mathcal{A}$ and $\mathcal{C}$ (if access binds instead of capacity, simply replace $\mathcal{C}$ with $\mathcal{A}$ 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 $\mathcal{T}$ types, the likelihood of drawing a type t is $(t/\mathcal{N}) = (t/\mathcal{T}t) = (1/\mathcal{T})$. The expected number of contacts by type is thus $(\mathcal{C}/\mathcal{T})$. Since agents reach their own knowledge bases with certainty, the expected knowledge profile of an agent i is $\mathcal{P}_i = \langle(\mathcal{C}/\mathcal{T})\kappa, (\mathcal{C}/\mathcal{T})\kappa, \ldots (1+\mathcal{C}/\mathcal{T})\kappa, \ldots (\mathcal{C}/\mathcal{T})\kappa\rangle$. Again using expected profiles so that vector elements are constants, we have for i and j in different communities $\mathrm{Cos}(\Theta_{ij}) =$

4.a

$$\frac{E[P_i] \bullet E[P_j]}{\|E[P_i]\| \|E[P_j]\|} = \frac{(T-2)\left(\frac{C}{T}\right)^2 \kappa^2 + 2\left(\frac{C}{T}\right)\left(1+\frac{C}{T}\right)\kappa^2}{(T-1)\left(\frac{C}{T}\right)^2 \kappa^2 + \left(1+\frac{C}{T}\right)^2 \kappa^2}$$

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

4.b

$$\frac{2\mathrm{C} + \mathrm{C}^2}{2\mathrm{C} + \mathrm{C}^2 + \mathrm{T}}$$

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 $\mathcal{X}$ of their channels to a specific type other than j and j's preference, then the expression for random association becomes

4.c

$$\frac{2(\mathrm{C}-\mathrm{X}) + (\mathrm{C}-\mathrm{X})^2}{2(\mathrm{C}-\mathrm{X}) + (\mathrm{C}-\mathrm{X})^2 + \mathrm{T}}$$

---

[3] We thank Paul Laskowski for this illustration.

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 see that narrower preferences increase distance $\mathcal{D}_{ij}$ between expected profiles, note that reserving channels creates profile $E[\mathcal{P}_i] = <((\mathcal{C}-\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}, ((\mathcal{C}-\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}, \ldots (1+\mathcal{X}+(\mathcal{C}-\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}, \ldots ((\mathcal{C}-\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}>$ with expected distance $\mathcal{D}(E[\mathcal{P}_i], E[\mathcal{P}_j] \mid t_i \neq t_j) = \sqrt{2}\ (1+\mathcal{X}+(\mathcal{C}-\mathcal{X})/\mathcal{T})\boldsymbol{\kappa}$ which increases in $\mathcal{X}$ for the type not targeted. □