The Simplicial Model of Social Aggregation, Information Diffusion, and Hyperlocal Human Dynamics on Social Media

Kerk F. Kee, Ph.D., Assistant Professor,
Chapman University, kerk.kee@gmail.com, www.ekerk.com

This position paper explores the 2015 SDSU HDMA workshop's research question, "What type of innovative research frameworks can help us to collect, analyze, visualize and predict hyperlocal human dynamics and social media?" Metzgar et al. (2011) maintain, "The term 'hyperlocal' brings to mind images of engaged citizens storming town halls seeking better governance and better reporting thereof" (p. 773). In fact, they define, "Hyperlocal media operations are geographically-based, community-oriented, original-news-reporting organizations indigenous to the web and intended to fill perceived gaps in coverage of an issue or region and to promote civic engagement" (p. 774). Kurpius et al. (2010) further state, "hyperlocal media operate at the crossroads of highly focused, locally-oriented news with technology-enabled potential as tools for civic engagement" (p. 360). In order for large-scale hyperlocal human activities to take place, Shirky (2009) suggests that members of a community need to know that everyone else in the community knows that everyone knows a community activity/effort is going to take place. In other words, information diffusion on social media is key to hyperlocal human dynamics. While the diffusion of innovations theory (Rogers, 2003) has tended to focus on the use of opinion leaders (OLs) with weak ties to disseminate new information, social network analysis literature documents that "stronger bonding ties may facilitate collective action" (Diani, 2011, p. 226), as in the case of the historical protest at Greensboro, NC, by four freshmen at North Carolina A. & T., a black college a short distance away (Gladwell, 2010). Therefore, this paper explores how geographically/locally based social groups and their social media connections can be tapped to diffuse and disseminate community oriented information, for public health interventions, political activism, environmental causes, etc. The focus is to model socially bonded groups with strong ties as mechanism for information diffusion and hyperlocal human dynamics.

In this paper, the *simplicial model of social aggregation* (SMSA) (Kee, Sparks, Struppa, Mannucci, & Damiano, in press; Kee, Sparks, Struppa, & Mannucci, 2013; Sparks, Kee, & Struppa, 2014) is introduced as a framework and approach for modeling bonded groups with 'shortened' social distance on social media. Based on the mathematical theory of simplicial complexes (see Faridi, 2002; Munkres, 1984), Kee et al. (2013) present a computational approach to modeling socially bonded groups with geometric spatial elements (mathematically called 'simplexes,' such as the full shaded triangles in Figure 1 in Appendix A) and a social aggregation a collection of simplexes and nodes (mathematically called a 'simplicial complex,' such as the entire network with nodes, linkages, and shaded triangles in Figure 1). In mathematical language, the faces of a 1-simplex are its two end points, the faces of a 2-simplex (a triangle) are its three edges, the faces of a 3-simplex (a pyramid) are its four faces (each of which is a triangle), and so on and so forth. Therefore, the face of an n-simplex is always an (n-1)-simplex. If a face is such that there is no other simplex in the complex containing it, it is called a maximal face or a facet. In Kee et al. (in press), the model further introduces the concept of a cover, *S*, of a simplicial complex as a set of nodes (or vertices) such that every maximal face in the complex contains at least one vertex of *S*. In other words, a cover is a set of individuals who encompass the complex in the following sense: every higher-dimensional group (i.e., a simplex), be it graphically an edge, a triangle, or a higher-dimensional face, contains *at least* one such vertex (or simplicial diffusers, SDs). A cover *S* is called *minimal* if no subset of *S* is also a cover. SDs are gatekeepers to their groups. Instead of traditional OLs with weak ties, the SMSA framework proposes to focus on SDs with strong ties as seeds for information diffusion.

Given the definitions of these terms, this position paper introduces the specific calculations of the information diffusion index (IDI) of individuals (such as OLs, SDs, and any nodes) in a social network. In Figure 1, the social aggregation (or simplicial complex) connected to D_2 shows all the pairwise connections and higher-dimensional groups (i.e., the shaded triangles) in the social network. Consider the simplicial complex whose nodes are D_2, x, y, and α, β, γ, δ, ε, θ. The reason Latin letters are used for some vertices and Greek letters for others will be evident shortly. For now, it will suffice to say that the vertices x, y are 'terminal' vertices, in the sense that they do not connect with any other vertex in the two simplexes (i.e., full triangles) but D_2, while this is not the case for the vertices with Greek letters.

From the classical network analysis point of view, D_2 is the OL (Valente & Pumpuang, 2007) because it has the highest degree centrality. However, the goal is to calculate the IDI of D_2 and compare it with the IDI of the pair {α, β} (i.e., mathematically a minimal cover, what this paper proposes to target for hyperlocal human dynamics and information diffusion on social media). In order to follow the calculations below, we will assign IDI's of 1 to any direct link (such as D_2-*x*, one degree separation), 0.5 to any double-link (such as D_2-*θ*-*α*, two degree separation), 1.4 (to approximate √2) to any simplicial link (such as *α*-*β*, because *α* and *β* belong to a socially bonded group represented by the full shaded triangle), 0.7 for simplicial double-link (such as *α*-*β*-*γ*), and 0.6 to a mixed double-link (such as D_2-*α*-*β*). The existence of a two-dimensional simplex (as in the triangle connecting *α*, *β*, and *γ*) shortens, so to speak, the distances between vertices, and increases the IDI's.

According to the SMSA framework, the IDI along a path of social distance *d* is $(1/2)^{(d-1)}$, and that an *n*-dimensional simplex has vertices whose social distance is 1/*n* (see Kee, et al., in press, for mathematical explanation). The IDI along a double-link is 0.5. However, along a simplicial double-link (such as *α*-*β*-*γ*), the distance from *α* to *β* is 0.5 because they belong to a full triangle (i.e., a simplex or a higher-dimensional group). The IDI of *α* on *β* is then $(1/2)^{(0.5-1)} = (1/2)^{(-0.50)} = 1.4$. Furthermore, the path *α*-*β*-*γ* is a simplicial double-link. Therefore, the IDI that *α* has on *γ* is 0.7, half of the direct flow from *α* to *β*. Since the mixed-second order links are in between a regular double-link and a simplicial double-link, the value is approximated to be the average between the two, which is 0.6. Furthermore, in keeping with a reasonable assumption by Christakis and Fowler's (2009), the calculation will consider up to second-order connections (i.e., double links; information diffusion from a node to his friend's friend). Therefore, for example, from the vertex D_2 we can reach *δ* and then *β*, but we will not consider further information diffusion that can be achieved by going from D_2 to *δ*, to *β*, to *ε*. With these two caveats in mind, the IDI's are 12.8 for *α*, 15.9 for *β*, 16.2 for D_2, and 28.7 for {*α*, *β*}. One could however argue that the average IDI for {*α*, *β*} is less than D_2 alone. Whether this is an important objection is debatable, because the set {*α*, *β*} still offers much better IDI than D_2 alone.

Furthermore, if we remove the vertices *x* and *y* as in Figure 2 the IDIs become 11.8 for *α*, 14.9 for *β*, only 14.2 for D_2, and 28.7 for {*α*, *β*}. Thus, despite the fact that D_2 remains the node with the largest degree centrality, we have now found that *β* alone has a higher IDI than D_2. Note also that the values for D_2 dropped by 2 (the number of terminal nodes removed), but 1 for *α* and *β* each, making *β* increasingly more powerful than D_2 in a situation where there are more simplexes than terminal vertices. Another worthwhile point is that the difference between the two methods (to target OLs with weak ties vs. SDs with strong ties) becomes larger and more visible as we increase the number of higher-dimensional simplexes in the network. Indeed, it is exactly when vertices belong to such higher-dimensional simplexes that the calculation will show increased IDI. So, for example, if *γ* and *ε* were also connected, and *β*, *γ*, *ε* make up the third full triangle, the IDI of *β* would grow again more than the one of D_2, thus making the simplicial method even more powerful. In this case, indeed, the IDI of D_2 would not change, while the IDI of *β* would grow by another √2 or 1.4, so that even in the case of the simplex with only one terminal vertex the average IDI of {*α*, *β*} would end up being higher than the IDI of D_2 alone.

The simplicial model of social aggregation (SMSA) (Kee, et al., in press; Kee, et al., 2013; Sparks, et al., 2014) has three implications for the 2015 HDMA workshop on "Studying Big Data and Social media with Hyperlocal Human Dynamics and Social Networks". First, the use of simplicial diffusing sets (i.e., the selected group of SDs who are gatekeepers to their groups with strong ties) to promote a health intervention, environmental cause, etc. takes advantage of the frequent internal communication patterns within socially bonded groups. Many often interact offline, as in the case of families, friends, roommates, work teams, etc. Furthermore, the simplicial approach can promote the positive manifestation of 'peer pressure' on members of bonded groups to attend and respond to the intervention message.

Second, individual OLs are respected by their peers, but these leaders are not likely to be family members, close friends, or roommates of a lot of people in their community. The simplicial approach takes the same communication load to be bore by a few OLs and share it among a group of SDs. The results could be more rapid diffusion. In a community with 100 members, a difference of 10 more simplicial diffusers instead of two or three opinion leaders in support of an initiative is a major difference. An active dissemination program (Dearing & Kee, 2012) can be further promoted.

Finally, Facebook or a similar social media platform can more easily be used for filtering users based on demographic and social indicators, such as location, gender, race, sexual orientation, etc., to help change agents better target specific at risk populations. Moreover, these messages are likely augmented by culturally embedded conversations within the target population, thus overcoming some of the linguistic and cultural barriers identified by Kreps and colleagues (2008).

Acknowledgement: This position paper is derived from collaborative work with Lisa Sparks, Daniele Struppa, Mirco Mannucci, and Alberto Damiano.

References

Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*. New York: Back Bay Books.

Dearing, J. W., & Kee, K. F. (2012). Historical roots of dissemination and implementation science. In R. C. Brownson, G. A. Colditz & E. K. Proctor (Eds.), *Dissemination and implementation research in health: Translating science to practice* (pp. 55-71). New York: Oxford University Press.

Diani, M. (2011). Social movemens and collective action. In J. Scott & P. J. Carrington (Eds.), *The SAGE handbook of social network analysis* (pp. 223-235). Thousand Oaks, CA: Sage.

Faridi, S. (2002). The facet ideal of a simplicial complex. *Manuscripta Mathematica, 109*, 159-174.

Gladwell, M. (2010, October 4). Small change: Why the revolution will not be tweeted. *The New Yorker, 4,* 42-49.

Kee, K. F., Sparks, L., Struppa, D. C., Mannucci, M., & Damiano, A. (in press). Information diffusion, Facebook clusters, and the simplicial model of social aggregation: A computational simulation of simplicial diffusers for community health interventions. *Health Communication*.

Kee, K. F., Sparks, L., Struppa, D. C., & Mannucci, M. A. (2013). Social groups, social media, and higher dimensional social structures: A simplicial model of social aggregation for computational communication research. *Communication Quarterly, 61*, 35-58. doi: 10.1080/01463373.2012.719566

Kreps, G. L., Neuhauser, L., Sparks, L., & Villagran, M. (2008). Translational community-based health communication interventions to promote cancer prevention and control for vulnerable audiences *Patient Education & Counseling, 71*, 315-350.

Kurpius, D. D., Metzgar, E. T., & Rowley, K. M. (2010). Sustaining hyperlocal media: In search of funding models. *Journalism Studies, 11*, 359-376.

Metzgar, E. T., Kurpius, D. D., & Rowley, K. M. (2011). Defining hyperlocal media: Proposing a framework for discussion. *New Media & Society*, 1461444810385095.

Munkres, J. R. (1984). *Elements of algebraic topology*. Menlo Park, CA: Addison-Wesley Publishing Company.

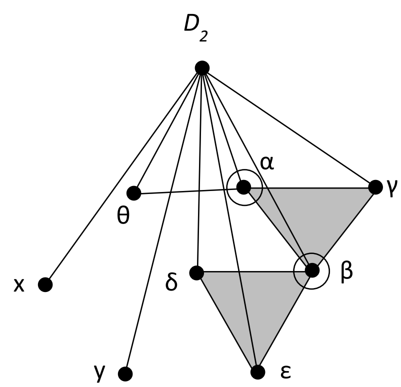
Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.

Shirky, C. (2009). *Here comes everybody: The power of organizing without organizations*. New York: Penguim

Sparks, L., Kee, K. F., & Struppa, D. C. (2014). Social aggregates. In T. L. Thompson (Ed.), *Sage encyclopedia of health communication, 3* (pp. 1272-1274). Thousand Oaks, CA: Sage.

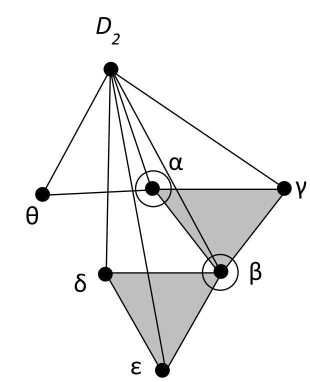
Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health Education & Behavior, 34*, 881-896.

Appendix A



16.2 for D_2, and 28.7 for {α,β}.

Figure 1



14.2 for D_2, and 28.7 for {α,β}.

Figure 1