

---

9-11-2015

## Information Diffusion, Facebook Clusters, and the Simplicial Model of Social Aggregation: A Computational Simulation of Simplicial Diffusers for Community Health Interventions

Kerk Kee

*Chapman University*, kee@chapman.edu

Lisa Sparks

*Chapman University*, sparks@chapman.edu

Daniele C. Struppa

*Chapman University*, struppa@chapman.edu

Mirco A. Manucci

*George Mason University*

Alberto Damiano

*Deledda International School*

Follow this and additional works at: [https://digitalcommons.chapman.edu/comm\\_articles](https://digitalcommons.chapman.edu/comm_articles)



Part of the [Communication Technology and New Media Commons](#), [Health Communication Commons](#), [Interpersonal and Small Group Communication Commons](#), [Leadership Studies Commons](#), [OS and Networks Commons](#), [Other Computer Sciences Commons](#), and the [Social Media Commons](#)

---

### Recommended Citation

Kee, K. F., Sparks, L., Struppa, D. C., Mannucci, M. A., & Damiano, A. (2016). Information diffusion, Facebook clusters, and the simplicial model of social aggregation: a computational simulation of simplicial diffusers for community health interventions. *Health Communication*, 31, 385-399. doi: 10.1080/10410236.2014.960061

This Article is brought to you for free and open access by the School of Communication at Chapman University Digital Commons. It has been accepted for inclusion in Communication Faculty Articles and Research by an authorized administrator of Chapman University Digital Commons. For more information, please contact [laughtin@chapman.edu](mailto:laughtin@chapman.edu).

---

# Information Diffusion, Facebook Clusters, and the Simplicial Model of Social Aggregation: A Computational Simulation of Simplicial Diffusers for Community Health Interventions

## Comments

This is an Accepted Manuscript of an article published in *Health Communication*, volume 31, issue 4, in 2016, available online at DOI: [10.1080/10410236.2014.960061](https://doi.org/10.1080/10410236.2014.960061).

## Copyright

Routledge/Taylor & Francis

**Information Diffusion, Facebook Clusters, and the Simplicial Model of Social Aggregation:****A Computational Simulation of Simplicial Diffusers for Community Health Interventions**

By

Kerk F. Kee, Ph.D. (corresponding author)

Assistant Professor

Department of Communication Studies

Chapman University

One University Drive

Orange, CA 92866

949-391-9988 (Google Voice)

714-532-6036 (Office)

714-628-7237 (Fax)

[kerk.kee@gmail.com](mailto:kerk.kee@gmail.com)[www.ekerk.com](http://www.ekerk.com)

Lisa Sparks, Ph.D.

Foster and Mary McGaw Endowed Professor in Behavioral Sciences Director, M.S. in Health and  
Strategic Communication Wilkinson College of Humanities & Social SciencesChapman University, One University Drive, Orange, CA 92866 Full Member, Chao Family  
Comprehensive Cancer Center/NCI Designated Adjunct Professor, Public Health/Medicine University of  
California, Irvine

Daniele C. Struppa, Ph.D.

Chancellor

Office of the Chancellor

Professor, School of Computational Science

Chapman University

Mirco A. Mannucci, Ph.D.

Department of Computer Science

George Mason University

&amp; HoloMathics, LLC

Alberto Damiano, Ph.D.

Mathematics - IBO Programme Deledda International School Corso Mentana 27 Genova, Italy

Please cite as:

Kee, K. F., Sparks, L., Struppa, D. C., Mannucci, M. A., & Damiano, A. (2016). Information diffusion, Facebook clusters, and the simplicial model of social aggregation: a computational simulation of simplicial diffusers for community health interventions. *Health Communication*, 31, 385-399. doi: 10.1080/10410236.2014.960061

Information Diffusion, Facebook Clusters, and the Simplicial Model of Social Aggregation:  
A Computational Simulation of Simplicial Diffusers for Community Health Interventions

Abstract

By integrating the simplicial model of social aggregation with existing research on opinion leadership and diffusion networks, this article introduces the constructs of *simplicial diffusers* (mathematically defined as nodes embedded in simplexes; a *simplex* is a socially bonded cluster) and *simplicial diffusing sets* (mathematically defined as *minimal covers* of a simplicial complex; a *simplicial complex* is a social aggregation in which socially bonded clusters are embedded) to propose a strategic approach for information diffusion of cancer screenings as a health intervention on Facebook for community cancer prevention and control. This approach is novel in its incorporation of interpersonally bonded clusters, culturally distinct subgroups, and different united social entities that co-exist within a larger community into a computational simulation to select sets of simplicial diffusers with the highest degree of information diffusion for health intervention dissemination. The unique contributions of the article also include seven propositions and five algorithmic steps for computationally modeling the simplicial model with Facebook data. (158 words)

**Authors' Key Words:** Simplicial model of social aggregation, higher-dimensional groups, theory of simplicial complexes, diffusion of innovations, opinion leadership, diffusion networks, and community health intervention.

**Acknowledgement:** In alphabetical order of last names, the authors would like to thank Bill Eadie, Susan Kung, Jake Liang, the editor and senior editor, and the two anonymous reviewers for their input and support of this project.

**Information Diffusion, Facebook Clusters, and the Simplicial Model of Social Aggregation:  
A Computational Simulation of Simplicial Diffusers for Community Health Interventions**

Health communication researchers have long been interested in the diffusion processes of innovations (i.e., new behaviors, ideas, objects, information, etc.) through communication networks for information diffusion, social influence, and behavioral change. In fact, this body of research has been interdisciplinary, with the goal of exploring how research can promote community health intervention, prevention, and control. For example, by combining Rogers' (2003, 2004) diffusion of innovations theory with Lewin's (1947) action research in social psychology and group dynamics, Wohlfeiler (1998) argues that understanding communication network processes can help spread a health intervention (as an innovation) to save lives, as in the case of the STOP AIDS project in San Francisco in the 1980s. STOP AIDS used the small group approach to recruit leaders to organize more small group meetings in order to diffuse safe sex practices in a self-sustaining fashion (Bertrand, 2004; Rogers, 2004; Singhal & Rogers, 2003). This approach can be extended to cancer prevention and control interventions via Facebook in today's changing communication landscape. Cancer prevention is a critical task because "it is better to prevent diseases than to concentrate resources on treating diseases after they become clinically apparent, when treatment may be too late to be effective" (Woolf, 2008, p. 2437).

With the rapid emergence of social media, such as Facebook, Twitter, and YouTube, the idea of fusing computer-mediated-communication literature, innovation diffusion theory, and communication network analysis for information dissemination is not only exciting and interesting, but also has much needed realistic applications. Lazer and colleagues (2009) argue that people engage in daily activities on the Internet and leave 'digital breadcrumbs', which "when pulled together, offer increasingly comprehensive pictures of both individuals and groups, with the potential of transforming our understanding of our lives, organizations, and societies in a fashion that was barely conceivable just a few years ago" (p. 721). Furthermore, today's information ecology is profoundly characterized by the active and participatory nature of social media users, especially in the health context (Chou, Hunt, Beckjord,

Moser, & Hesse, 2009; Chou, Prestin, Lyons, & Wen, 2013; Eysenbach, 2008).

According to Chou and colleagues (2009), the Internet and social media can aid in health promotion efforts in three ways. First, these web-based platforms can provide an increased sense of social support and connection among users (Idriss, Kvedar, & Watson, 2009; Wangberg et al., 2008). Second, these patient driven platforms can promote information sharing (Eysenbach, 2008) that is user-centered (Hawn, 2009). Third, social media can expand the reach of traditional health promotion efforts through online social networks (Thackeray, Neiger, Hanson, & McKenzie, 2008; Vance, Howe, & Dellavalle, 2009), including at the global level (Norman, McIntosh, Selby, & Eysenbach, 2008). In fact, Chou and colleagues refer to social media as dissemination channels that can be effectively leveraged for health communication campaigns. We believe by computationally modeling communication interactions captured by Facebook and similar platforms, a change agent can strategically target individuals in different bonded social clusters as seeds to further disseminate the information within their clusters and accelerate a community health intervention.

We explore in this article *how the participatory nature of Facebook can be leveraged to improve cancer prevention and control*. More specifically, we pose the research question: *How can socially bonded clusters be mathematically modeled on Facebook to strategically accelerate peer-to-peer information diffusion for community health interventions?* In order to answer this question, the article is organized into six sections. First, we review the literature on opinion leadership and innovation diffusion while introducing the constructs of higher-dimensional groups (i.e., socially bonded clusters) and social aggregations (i.e., a complex network of individuals and socially bonded clusters along with their social ties) to set the stage for a multi-dimensional picture of communication networks. Second, we present seven propositions to describe how information diffusion, higher-dimensional groups, social aggregations, and simplicial diffusing sets on Facebook can be mathematically conceptualized, thus extending classical network analysis to higher dimensionality. But briefly here, the term ‘simplicial’ is used to describe higher-dimensional groups and/or socially bonded clusters mathematically defined as a

‘simplex’ (more explanation on p. 7). Third, we elaborate on how the simplicial model extends similar concepts and theories in classical network literature to higher dimensionality. Fourth, the readers are guided through five algorithmic steps for applying the seven propositions to higher-dimensional groups and social aggregations. Fifth, we demonstrate through an example simulation on how the propositions and algorithmic steps can be applied to identify and select sets of simplicial diffusers (i.e., individuals who are strategically poised to diffuse information to their social clusters) for community health interventions. Finally, implications and limitations are discussed in the conclusion.

### **Opinion Leaders, Innovation Diffusion, Higher-Dimensional Groups, and Social Aggregations**

The strategy of using opinion leaders to accelerate health intervention diffusion has received much attention in public health (Kelly et al., 1991; Valente & Pumpuang, 2007). In the health context, an opinion leader can be understood as the person who influences others to adopt a new health related object, behavior, practice, etc. Rogers (2003) describes an opinion leader as someone who is respected for their opinions about an innovation, something perceived as new by potential adopters. Kadushin (2012) notes that an opinion leader usually represents the norms of the community, thus s/he is relatable to most members within the community. Bertrand (2004) explains that community members often respect an opinion leader’s knowledge about an innovation and their reputation within the community. Furthermore, Valente (1995) describes opinion leaders as the few individuals in the network who have the largest number of social ties. In short, individuals whose opinions, knowledge, and reputation are respected by and relatable to community members and who have many connections are well-positioned to diffuse information or influence the opinion of a given health innovation within a community, either for or against it.

Katz and Lazarsfeld (1955) propose a model that describes the flow of ideas from mass media through influential opinion leaders to the masses connected to the opinion leaders. In fact, “studies of diffusion and mass communication have shown that people rarely act on mass-media information unless it is also transmitted through personal ties” (Granovetter, 1973, p. 1374). Katz and Lazarsfeld’s model was

revolutionary during its first proposal because the argument challenged the common assumption that mass media has a direct influence on the masses, and highlighted the dissemination role of interpersonal communication in the diffusion process. Subsequently, network analysis based on dyadic ties have been utilized extensively by researchers to identify opinion leaders within larger communities (Valente & Davis, 1999) and used as a means for gaining further understanding of how new ideas, behaviors, objects, and information spread through various channels (Dearing, 2004; Rice, 2001; Rogers, 2004). Practically, Southwell (2013) explains that opinion leaders are regarded as ‘information hubs,’ therefore, they are often recruited by strategists to endorse certain messages and to drive message diffusion in target social networks.

However, Southwell (2013) also maintains, “purely egalitarian connection networks are not the rule; instead, we tend to see clustering.” In fact, some clustering is based on interpersonal bonding. Kee, Sparks, Struppa, and Mannucci (2013) maintain that within a larger community, there are interpersonally bonded clusters of individuals, culturally distinct subgroups within a society, and different united social entities that co-exist within a system. Examples of such groups can include families, friendship groups, work teams, cultural/immigrant groups, political groups, religious groups, etc. In other words, members of these groups share a high degree of what Putnam (2000) terms bonding social capital. Zubcsek and colleagues (2014) refer to these as “cohesive subsets in social networks” (p.50). Furthermore, the notion of culturally distinct subgroups is important because Southwell (2013) maintains that there is usually homogeneity in conversational networks regarding critical health topics.

According to Kee and colleagues (2013), members of socially bonded clusters, referred to as *higher-dimensional groups*, do exist and they often shared three interrelated properties and uniting conditions: (a) a private sphere and world view, a sense of shared identity and ideology, and objectivation of reality, (b) an open in-group communication climate with a high degree of trust or a converging climate of conformity and cohesion, and (c) a shared life story, narrative history, a coherent past, and an anticipated future together. These properties cannot be simply reduced to the dyadic ties that constitute



the bonded group. These groups exist within a *social aggregation*, a system (or an organization, a community, etc.) that is made up of a collection of individuals, together with their mutual connections (partnerships, friendships, relations, etc.), as well as the collection of possible higher-dimensional groups within the system. Therefore, a higher-dimensional group within a social aggregation is a collection of clustered individuals that evolve into inextricably tied entities with distinct uniting conditions. Recently, Zubcsek and colleagues (2014) suggest that when cohesive subgroups overlap, an information community emerges to facilitate efficient communication they defined as lossless information transmission. Cohesive groups regulate information flow, and these subgroups are spaces where information exchange is more likely to take place.

Social aggregations and its higher-dimensional groups are important for community health interventions because as Kadushin (2012) argues, “cohesive network[s]” (p. 147), “[g]roups, [and] circles ... may be more effective in social influence and diffusion than single individuals” (p. 148). He further argues, “conceptualizing diffusion as taking place through circle or focused-based clusters is a more fruitful way to proceed” (p. 161). However, information flow, opinion diffusion, and social influence are conceptually distinct but methodologically difficult to distinguish empirically (Weimann, 1994). Therefore, we limit the scope of the mathematical conceptualization of the simplicial approach to information diffusion, and reserve social influence for future research.

Zubcsek and colleagues (2014) argue that for efficient communication to occur, the most stable cohesive groups need to share a high degree of connectivity. While they simply operationalized connectivity as frequency of email communication and duration of telephone calls, Kee and colleagues (2013, p. 47) advance that a higher-dimensional group can be detected by the following activities, providing a picture of how high connectivity among bonded clusters can manifest on Facebook:

1. “The expression of a private sphere and world view, a shared identity and ideology, and objectivation of reality can be seen in overlapping profile information, joining and participating in overlapping Facebook groups, endorsing (e.g., clicking the ‘Like’ button [for]) the same sets of postings, using similar language, code words, and linguistic styles, etc.”
2. “The presence of an open communication climate or a convergence climate can be demonstrated by members interacting with each other on Facebook via multiple channels, including publicly

commenting on each other's status updates and postings, sending each other private individual and group messages, chatting in real time, etc. Furthermore, we may also observe the use of emoticons to show emotional intimacy and intensity."

3. "A shared life story, narrative history, a coherent past, and an anticipated future together could be documented by tagging each other in common pictures, videos, and postings, and indicating the intention to attend common offline events organized and publicized on Facebook, etc."

Given the arguments above, we integrate the constructs of opinion leaders, higher-dimensional groups, and social aggregation to propose an innovative approach for accelerating health innovation diffusion and dissemination through tapping *simplicial diffusers* on Facebook. In addition to individuals who can diffuse information and potentially influence a community, we also look for the smallest groups of individuals (i.e., a simplicial diffusing set) who can collectively play that role. Our goal is to show a complementary approach to the use of individual opinion leaders. In the example simulation, we demonstrate a scenario of targeting three opinion leaders and six simplicial diffusers for innovation diffusion. In the next section we will present seven propositions of the *simplicial model of social aggregation* and explain the mathematical foundation underlying the proposed approach.

### **Seven Propositions of the Simplicial Model of Social Aggregation**

The simplicial model of social aggregation (Kee, et al., 2013; Mannucci, Sparks, & Struppa, 2006; Sparks, Kee, & Struppa, 2014) is a theoretical framework applying the mathematical theory of simplicial complexes (see Beck, 1979; Faridi, 2002; Munkres, 1984) to studying higher-dimensional groups and social aggregations. The use of the simplicial approach in the social sciences was first introduced by Mannucci, Sparks, and Struppa (2006) and recently employed by Ren and colleagues (2011) to model wireless communication networks and cited by Brooke-Taylor and Testa (2013) in a study of infinite random simplicial complex. The model was first proposed by Mannucci et al. (2006) in a paper archived at Cornell's e-print service (<http://arxiv.org>). Kee et al. (2013) theoretically explicated the model for communication research, and conceptualized bonded clusters as higher-dimensional groups (mathematically defined as *simplexes*) and a complex community as social aggregation (mathematically defined as a *simplicial complex*) in which bonded clusters are embedded. Most recently, Sparks et al. (2014) advanced the model to the health communication context, arguing that the simplicial approach

allows researchers to better capture the complexity of social connections for more strategic health campaigns. The unique contributions of the present article is threefold: the mathematical conceptualization of *simplicial diffusers* as nodes within simplexes and a *simplicial diffusing set* as a minimal cover of a simplicial complex; the quantitative measure of *information diffusion* to rank simplicial diffusing sets; and a series of methodological propositions and simulation steps for computationally identifying clusters diffusers and simplicial diffusing sets on Facebook. We explain these mathematical concepts next.

According to Struppa, Sparks, Mannucci, and Kee (2006; 2013; 2014), in mathematics the term ‘simplicial’ is used to describe a geometric spatial element (also known as a ‘simplex’) with the minimum number of boundary points, such as a one-dimensional line segment, a two-dimensional triangle, a three-dimensional pyramid, etc. In this article, geometric spatial elements are used to represent *higher-dimensional groups*, such as interpersonally bonded clusters, culturally distinct subgroups, united social entities, and other inextricably tied social groups and socially cohesive clusters in a communication network of social aggregation. In turn, a *social aggregation* is mathematically represented by a simplicial complex, which is a complex network of multi-dimensional simplices, including nodes, dyadic linkages, and higher-dimensional groups as simplices in the forms of full triangles, pyramids, etc. Therefore, the simplicial model extends classical network analysis to higher-dimensionality.

We further elaborate on the mathematical foundation below, using a breast cancer screening message and Facebook as a case in point. We choose Facebook as our simulation platform because Chua, Madej, and Wellman (2011) argue that Facebook has become an integral part of many people’s daily communication repertoire: “People continue to be social when they are indoors. They chat online with friends, keep up with them via their Facebook pages, ... meet them offline to round out discussions, and then meet online again to talk about other things” (p. 103). Furthermore, it is a common practice that people spent their free time perusing their friends’ Facebook profiles. Given Facebook’s highly diffused status and heavy daily usage, we believe it is one of the best platforms for understanding how the

simplicial model can be used for health interventions. The propositions below also extend Struppa, Sparks, Mannucci, and Kee's (2006; 2013; 2014) theoretical model to methodological procedures introduced in this article.

**Proposition 1:** *Two individuals who are 'friends' on Facebook are capable of spreading a health innovation to each other in a symmetrical way.*

If Alicia (A) and Bhayana (B) are Facebook friends, Alicia is assumed to have as much potential to diffuse a piece of information to Bhayana as Bhayana has to do likewise to Alicia through passing along, let's say a YouTube video about, a breast cancer screening intervention message via Facebook. One way to observe such information diffusion is through a scenario of YouTube video on Facebook. More specifically, Alicia clicks the 'Like' button and posts the video on her Facebook page, and Bhayana watches the same video because of Alicia's posting. At the end of the video, viewers are called upon to click the 'Like' button and re-post the video on their Facebook page. Consequently, Bhayana also clicks the 'Like' button and posts the YouTube video on her Facebook page. When this pairwise spreading of an intervention message expands from a single source to a large network of people, an intervention is said to have become 'viral' (through self-replicating processes, similar to the spread of a virus in the medical sense).

**Proposition 2:** *Indirect information diffusion can be wielded, and its reach decreases exponentially with the number of intermediaries.*

If Alicia is a friend of Bhayana on Facebook, and Bhayana is a friend of Chandra (C) on Facebook, then Alicia can diffuse a breast cancer screening intervention YouTube video to Chandra through Bhayana. Note that in this case, Alicia is not directly connected to Chandra online and offline. Therefore, this is not an assumption of transitivity. This second-order information diffusion, however, is assumed to be 1/2 or 50% as effective, strong, or likely as a direct information flow or diffusion. If four people are indirectly connected on Facebook (Alicia, to Bhayana, to Chandra, to Dyani), then the information diffusion of the YouTube video via Facebook will be computed as 25% of a direct flow.

Because Borgatti and Lopez-Kidwell (2011) argue, “Flows matter in most network theories but are generally assumed immeasurable in practice” (p. 45) and that network scholars “build theory that links the observable network of social relations to these latent flows” (p. 45), we simply offer in this article an example of how one could treat the change in information diffusion as the nodes get more distant. We can work with a model based on any assumption. The choice of 50% is arbitrary. Therefore, the quantitative measurement is susceptible to refinement through empirical and experimental data in future research.

In simulation research, one may want to have a model in which information diffusion propagates only up to some maximum number of steps. This will have significant computational implications, which will make the algorithms proposed later in this article simpler and faster. This is a reasonable assumption because in most social and health community contexts, it is unlikely that Alicia can reach Reisha through a chain of, say, ten friends (except in a ‘viral’ case). Indeed, with the assumption of ‘exponential decay’ (Provencher, 1976), information diffusion through the network, numbers quickly become small. Therefore, we use Christakis and Fowler’s (2009) argument to limit information diffusion potential that Alicia has on somebody who is more than three steps away from her on Facebook can be assumed to be basically negligible. As Christakis and Fowler discuss in an example of three piano teachers in Tempe, AZ, the word-of-mouth network of recommendations fizzle out after three degrees of separation (i.e., can be assumed to approach zero with “the teacher’s friend’s friend’s friend”). However, in a simulation, the values can be set by the researchers based on any and various assumptions.

**Proposition 3:** *Information diffusion between two individuals is quantitatively the same for all dyadic (or pairwise) friendship structures on Facebook.*

While some friendships are qualitatively stronger than others, in order to simplify for computational simulation, it is reasonable to assume that all pairwise friendship structures on Facebook are quantitatively equivalent, and incorporate differences in terms of friendship strengths in future research.

**Proposition 4:** *At the interpersonal level, the degree of information diffusion that A has on B on*

*Facebook is the sum of all the direct and indirect information diffusion from A to B.*

In other words, the degree of information diffusion Alicia has on Bhayana can be computed by looking at all the possible ways through which Alicia can reach Bhayana (maybe directly as friends, and indirectly as friends of friends, etc.) and then adding up all the degrees of information diffusion along all possible paths connecting Alicia to Bhayana. The assumption that if Alicia is connected to Bhayana through multiple friends, then Alicia would have a greater degree of information diffusion on Bhayana (thus more likely to get the same information to Bhayana and trigger Bhayana to further diffuse the information) is based on the mere exposure effect (Unkelbach, Fiedler, & Freytag, 2007). Daly (2011) further explains this strategy in his book on advocacy, “the more often people see an object, the more positive they feel about it”. More recently, Mayer-Schönberger and Cukier (2013) document, “a big predictor that people would take an action (post content, click an icon, and so on) was whether they had seen their friends do the same thing” (p. 114). In other words, if Alicia posts a YouTube video on Facebook, and a common friend (or more) between Alicia and Bhayana also post(s) the video, Bhayana will be more likely to act on the message, due to increase liking resulting from mere repetitive exposure and seeing their friends’ behaviors.

**Proposition 5:** *At the community level, the global degree of information diffusion that an individual has on the community is the sum of the degrees of information diffusion the individual has on separate individuals.*

Thus, for Alicia in the community, her global degree of information diffusion  $i(A)$  can be defined as the sum of the degrees of information diffusion she has on all the individuals in the community on Facebook. For example, let’s assume a health intervention message from Alicia can diffuse to Bhayana in three different ways on Facebook, assuming that they are not Facebook friends directly: through a path of two linkages (i.e., a second-order connection or a double-link; with one intermediary), through a path of three linkages (with two intermediaries), and through a path of four linkages (three intermediaries). As previously explained in Proposition 2, the degree of information diffusion between two individuals with

one intermediary is  $1/2$  (or 50%) as effective as a direct information flow between them. Therefore, the degree of information diffusion along the first path in this scenario is  $1/2$  (or 50%) of a direct flow, the second is  $1/4$  (or 25%) of a direct flow, and the third is  $1/8$  (or 12.5%) of a direct flow. Note: the degree of information diffusion along a path is  $(1/2)^{(d-1)}$ , where  $d$  is the distance between two nodes. The distance of a second-order connection or a double-link is '2'. This distance is later referred to as *social distance* or *simplicial distance* in Proposition 6.

If one follows Christakis and Fowler's (2009)'s argument cited in Proposition 2, the information diffusion through three intermediaries can be assumed to be zero instead of 12.5% of a direct flow. However, we allow this illustration to remain the 12.5% value to show that how the equation of  $(1/2)^{(d-1)}$  can be used and that values for a simulation study are determined by the researchers during each attempt. As a result, the total degree of information diffusion from Alicia to Bhayana via Facebook, in the spread of a breast cancer screening intervention YouTube video, is equal to  $i=1/2+1/4+1/8=7/8$ .

However, an additional complexity needs to be introduced in the picture: the simplicial model described in this article uses not only nodes and linkages (or edges in mathematical language; individual actors and social ties in classical network analysis literature), but also higher-dimensional groups. According to the *theory of simplicial complexes* (see Beck, 1979; Faridi, 2002; Munkres, 1984), and further explicated by Struppa, Sparks, Mannucci, and Kee (2006; 2013; 2014) for social science research, isolated nodes are 0-dimensional structures. When some (or all) of the nodes are connected, a structure that is 1-dimensional (though it may still have 0-dimensional components) is produced. In mathematics, nodes are known as 0-simplices while linkages are known as 1-simplices. When nodes with linkages are connected, a simplicial complex (i.e., a complex consisting of simplices with 0- and 1-dimensional structures in this case) is produced.

Moving to a higher dimension, a triangle (to be more precise, a full triangle, which also includes its interior and is, therefore, two-dimensional) is known as a 2-simplex. In this article, a higher-dimensional group of three friends [A,B,C] on Facebook can be mathematically modeled with a 2-

simplex, as shown in Figure 1 in Appendix A. In mathematical language, A, B, and C are vertices of the triangle. Note that a triangle also has three one-dimensional faces or 1-simplices (i.e., sides, edges, or segments in mathematical language; linkages in classical network literature). In turn, every 1-simplex has two faces or 0-simplices (i.e., end points or border points in mathematical language; nodes in classical network literature). The faces of a 1-simplex are its two end points, the faces of a 2-simplex (a triangle) are its three sides, 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*.

We previously explained that when nodes with linkages are connected, a simplicial complex (i.e., a complex consisting of simplices with 0- and 1-dimensional structures) is produced. This argument can be extended further: when nodes with linkages and triangles are connected, another simplicial complex is produced, but this complex consists of 0-, 1-, and 2-dimensional structures. Therefore, simplicial complex can be defined as the union of several simplices (possibly of different dimensions), including all their faces. A *simplicial complex* is a complex network of multi-dimensional simplices, including nodes, dyadic linkages, and higher-dimensional groups as simplices in the forms of full triangles, pyramids, etc. On Facebook, a simplicial complex represents a multi-dimensional social aggregation with individual nodes, higher-dimensional groups, and the social ties among the members, all via digital connections.

Returning to the attempt in this article to extend prior work on the simplicial model, recall the argument advanced in Proposition 5: the total degree of information diffusion from A to B via Facebook can be calculated as  $i=1/2+1/4+1/8=7/8$ . What is the effect of a 2-simplex (or  $n$ -simplex) on this calculation? The *social distance* between two individuals belonging to a higher-dimensional simplex is assumed to be shorter than if they are not in a higher-dimensional group because, in some sense, the simplex is a closed entity itself; it exists ‘as a group’. Sabidussi (1966) introduced the concept of *closeness centrality* as the inverse of the average distance of a vertex from all the other vertices in a square. In this article, the *degree of information diffusion* is defined as a generalization to higher



dimensions of what can be done with graphs in classical network theory, similar to Sabidussi's *closeness* centrality. Through this notion of information diffusion, the concept of closeness (or social distance, reciprocal distance, simplicial distance) is made available for simplicial complexes also. Therefore, another proposition is introduced below:

**Proposition 6:** *In a higher-dimensional group defined as an  $n$ -dimensional simplex, the social distance (or simplicial distance) between two of its members is  $1/n$  or the smallest possible  $1/n$  if the simplex belongs to more than one facet of the simplicial complex.*

Let us assume that a full triangle and a full pyramid share a common edge. In the triangle [B,C,E] (i.e., a higher-dimensional group on Facebook), the social distance (or simplicial distance) between any two of its vertices can be assumed to be  $1/2$  of the regular pairwise connection (i.e., as in a 1-simplex or simply a linkage). This simplicial distance is a measure of social distance between two nodes or connectedness among a set of nodes. In the pyramid [A,B,C,D] (i.e., another higher-dimensional group on Facebook), the distance between any two vertices is assumed to be  $1/3$  of a regular pairwise connection. Proposition 6 states that when a linkage (or an edge) is part of more than one facet of the simplicial complex, such as in the case of the shared edge between B and C because B and C belong to both [B,C,E] and [A,B,C,D], the social distance between the two nodes is assumed to be the minimum possible ( $1/3$  in this case). Note that Propositions 4 and 5, the two additivity assumptions, offer a reasonable approximation of the reality of information diffusion on Facebook. The more ways one can reach an individual on Facebook, the more likely one is able to diffuse a health intervention message to her.

**Proposition 7:** *A simplicial diffusing set can be defined as a minimal collection of individuals who can reach out to the entire community or social aggregation (through their clusters).*

A *simplicial diffusing set* is a generalization for *groups* similar to the concept of *individual opinion leaders* discussed earlier. As previously explained, the STOP AIDS intervention used the small group approach to recruit leaders to organize more small group meetings to diffuse safe sex practices in a self-sustaining fashion. Therefore, instead of individuals with such skills as described, we look for the

smallest groups of individuals who can *collectively* play that role.

Given the notions of a simplicial complex and maximal face discussed in Proposition 5, the simplicial model looks for a set of individuals who can ‘span’ the complex (i.e., a subset of ‘boundary spanners’ embedded within simplexes of the entire simplicial complex). This set of simplicial diffusers is called a *cover* of the simplicial complex. From a mathematical point of view, a cover,  $S$ , of a simplicial complex is defined to be 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, maximal face, or facet), be it graphically an edge, a triangle, or a higher-dimensional face, contains *at least* one such simplicial diffusers. A cover  $S$  is called *minimal* if no subset of  $S$  is also a cover.

In Figure 1, the example simplicial complex is made up of facets ABC, CD, and BD, which are higher-dimensional groups. The sets  $\{A,C,D\}$  and  $\{C,D\}$  are both covers, because they connect the three separable facets together as a simplicial complex. However, only the set  $\{C,D\}$  is minimal between the two. This minimal cover is indicated with an asterisk \* on the corresponding nodes.  $\{C,D\}$  is identified as one of the minimal covers. We will later discuss how to identify the optimal cover based on degrees of *information diffusion* (see Table 1). Besides  $\{C,D\}$ , other possible covers are  $\{A,D\}$ ,  $\{B,C\}$ , and  $\{B,D\}$ , which are also minimal covers. Because the goal is to identify minimal covers for health intervention diffusion, any of the four facets could be selected for the simplicial complex in Figure 1. The non-minimal covers would be  $\{A,B,C,D\}$ ,  $\{A,C,D\}$ ,  $\{A,B,D\}$ ,  $\{B,C,D\}$ , and  $\{A,B,C\}$ .

Given the discussion above, a *simplicial diffusing set* can be mathematically defined as a set of individuals belonging to a *minimal cover* of the complex. If a message spreads from the members of a simplicial diffusing set, it will rapidly reach the entire complex, since each facet of the complex contains at least one of those individuals. As noted before, other possible minimal covers are  $\{A,D\}$ ,  $\{B,C\}$ , and  $\{B,D\}$ ; therefore, they are possible simplicial diffusing sets for the complex in Figure 1.

### **Simplicial Extension of Classical Network Concepts and Theories based on Pairwise Connections**

At this point in the article, one might wonder in what ways is the simplicial approach extending existing literature? This section will compare and contrast the proposed concepts with a few well-established network concepts in the classical network literature, such as bridges, brokers, strength of weak ties, structural holes, and existing algorithms for identifying social clusters.

*Bridges/Brokers vs. Simplicial Diffusers.* A similar concept is a ‘bridge,’ defined as “a line such that the graph containing the line has fewer components than the subgraph that is obtained after the line is removed (nodes incident with the line remain in the subgraph)” (Wasserman & Faust, 1994, p. 114). These nodes are also referred to as ‘bridges’ (Monge & Contractor, 2003). Mische (2011) describes ‘bridges’ as “network intersections” (p. 86). Marin and Wellman (2011) explain, “bridging is a structural condition where the tie creates a connection between previously unconnected portions of a network” (p. 14).

Another similar concept is ‘broker,’ defined as an actor who “falls on the paths between others” (Hanneman & Riddle, 2011, p. 359). Gould and Fernandez (1989) explicate five types of brokers: A *coordinator* is an actor who connects two other in-group members (i.e., the coordinator belongs to the same group as these two members). A *consultant* is an actor who connects two other out-group members who belong to the same group (but the consultant does not belong to the group in this case). A *gatekeeper* is an actor who connects an out-group member with an in-group member (the gatekeeper belongs to the same group as the in-group member). A *representative* is the reverse of the gatekeeper, where an actor connects an in-group member from his/her own group to an out-group member. A *liaison* connects members of two different groups, neither of which is his/her own group.

Based on Gould and Fernandez’s (1989) typology, the notion of simplicial diffusers is more like a gatekeeper. However, a simplicial diffuser does not simply belong to a group, but a higher-dimensional group (with uniting conditions discussed on p. 5, so the tightly bound group is mathematically defined as a simplex). For that reason, simplicial diffusers represent a unique subset of brokers in classical network literature, one that represents the ‘gatekeepers’ to tightly bonded social groups. If there are existing

connections between a change agent and these simplicial diffusers, then they will be considered ‘bridges’ to the corresponding higher-dimensional groups.

*Strength of Weak Ties/Structural Holes vs. Simplicial Model of Social Aggregation.* Two other well-established theories are relevant to the discussion at hand: Granovetter’s (1973, 1983) strength of weak ties and Burt’s (1992) structural holes. A common theme between these two theories is the notion of information diffusion through weak ties and nonredundancy between contacts. First, Granovetter’s theory stresses “the strength of dyadic ties” (p. 1360) and that the strength of an interpersonal tie is “linear” (p. 1361) between the two partners of a dyad, based on time spent, emotional intensity, intimacy, and reciprocal services. He argues that weak ties’s strength is its ability to diffuse information from one fragmented portion of a network to another, often through bridges (which can only be weak ties).

In the second theory, Burt (1992) defines a structural hole as “the separation between nonredundant contacts... A structural hole is a relationship of nonredundancy between two contacts” (p. 18). In other words, structural holes are lack of connectivity. Brokers, filling structural holes, connect these contacts. In Burt’s conceptualization, “As a result of the hole between them, the two contacts provide network benefits that are in some degree additive rather than overlapping” (p. 18). As most people develop relations with others who are like themselves, the same pool of information tends to recycle within a social group that emerges out of homophily. Therefore, when a person bridges structural holes between groups, s/he receives opportunities for information access otherwise not accessible, being informed of an opportunity early or at a better timing than others, being referred to future opportunities by people in different networks.

There are three ways in which the simplicial model extends Granovetter’s and Burt’s theories. First, in the two classical theories, a social group is held together by one-dimensional pairwise connections (i.e., a line connecting two nodes). The simplicial model introduces the notion of a higher-dimensional group (defined as a simplex) to describe socially bonded clusters, and the notion of a social aggregation (defined as a simplicial complex) to provide a multidimensional representation of the

clustering and fragmentation the two classical theories describe with one-dimensional social ties. Furthermore, a socially bonded cluster or a higher-dimensional group on Facebook are not simply any social groups, they manifest the uniting conditions described on p. 5 that make these groups uniquely identifiable on social media.

Second, weak ties and structural holes are strategic opportunities for new information to spread across social groups, but these are naturally existing ties and holes in a given social network. For example, the traditional approach identifies traditional opinion leaders who have the largest numbers of ties, although these are mostly weak ties) in a given social system, and partner with them to promote a health intervention to the rest of the social network. A potential limitation is the possibility of overlooking isolated groups that are not connected to the opinion leaders within a social system. The simplicial model describes a strategy that goes beyond these traditional opinion leaders, and focuses on the higher-dimensional groups. Then a change agent identifies the smallest set of individuals who are gatekeepers of all the higher-dimensional groups embedded within a social system. The set is called a ‘simplicial diffusing set’, defined as a ‘minimal cover’ of a simplicial complex. Then a change agent actively reaches out to the simplicial diffusing set to promote a health intervention to the rest of the social network through their respective clusters.

Third, in contrast to the opinion leadership approach which takes advantage of weak ties, the simplicial approach taps into simplicial diffusers’ membership in higher-dimensional groups because “stronger bonding ties may facilitate collective action” (Diani, 2011, p. 226). In fact, Gladwell (2010) argues, “weak ties seldom lead to high-risk activism”. In the opening story of his article, he tells the historical protest at Greensboro, NC, by four freshmen at North Carolina A. & T., a black college a short distance away. He explains, “So one crucial fact about the four freshmen at the Greensboro lunch counter—David Richmond, Franklin McCain, Ezell Blair, and Joseph McNeil—was their relationship with one another”. Also, McNeil and Blair were roommates in a campus dormitory, and Richmond and McCain shared a room just one floor up. Moreover, Blair, Richmond, and McCain went to the same high

school. He continues, “The four would smuggle beer into the dorm and talk late into the night in Blair and McNeil’s room.” What Gladwell describes is an example of a higher-dimensional group in the simplicial model.

The simplicial approach to utilize strong ties for intervention diffusion within bonded cluster is aligned with existing network literature. For example, Burt (1992) suggests, “the likelihood that information will move from one person to another is proportional to the strength of their relationship” (p. 19). Furthermore, van der Hulst (2011) maintains, “Dense cohesive networks facilitate coordination within the group, increase group compliance” (p. 259) and “people are also more likely to be influenced by their close friends and associates” (p. 259). Carrington (2011) explains, “Strong ties, such as family and close friends, tend to know one another and therefore tend to form closed communication circles, in which the same information and attitudes are recycled” (p. 239). The simplicial model takes advantage of information recycling within bonded groups to diffuse a health intervention.

*Existing Cluster Algorithm vs. Simplicial Algorithm.* It is true that in recent years many different algorithms have been devised to uncover groups within networks of all kind, and therefore also within social networks. Among those one should certainly consider simulated annealing (Boyd, 1991; Dekker, 2001), multidimensional scaling (Arabie & Carroll, 1989; Carroll & Arabie, 1998; Freeman, Romney, & Freeman, 1987), hierarchical clustering (Breiger, Boorman, & Arabie, 1975), and a few other summarized by Freeman (2011). Furthermore, Zubcsek and colleagues (2014) also review several models of cohesive groups, such as Lambda sets (Borgatti, Everett, & Shirey, 1990), UCINET clique clustering (Everett & Borgatti, 1998), and uniform communities (Palla, Derényi, Farkas, & Vicsek, 2005). However, it should be pointed out that virtually all the algorithms and models that are currently used originate from various engineering contexts, and employ substantially different mathematical tools.

For example, simulated annealing (whose name betrays the origin from metallurgic problems) is essentially based on minimization of the energy of different states of the system, and then using a probabilistic approach to decide whether or not the system should move to a different state.

Multidimensional scaling, on the other hand, uses the relatively new ideas from multiscale approximation embedded in the notion of wavelets, and thus is essentially a form of Fourier analysis. Hierarchical clustering is based on a notion of metric (or several notions of metric) in the space of configurations, and so on. The simplicial approach, on the other hand, introduces a different point of view, by only exploiting the topological structure of the network, as a higher-dimensional simplex. Thus no metric is necessary, nor we use any tools (say) from Fourier analysis. As such, we bring a different kind of mathematics to bear to the problem. The article now turns to a discussion on five algorithmic modeling steps.

### **Algorithmic Steps for Computational Modeling of Social Aggregation on Facebook**

Given the seven propositions and mathematical foundation of the constructs of information diffusion, higher-dimensional groups, social aggregations, and simplicial diffusing sets, we show how the simplicial model of social aggregation can be applied to help provide strategic recommendations to change agents for community health interventions. In this section, we extend Wohlfeiler's (1998) STOP AIDS example in the 1980s in San Francisco to a hypothetical scenario on Facebook in the 2010s in Los Angeles for cancer prevention and control, a campaign we call CANCER SCREENING. Woolf (2008) reports that screening can reduce colorectal and breast cancer mortality by 15% and 20% respectively. The simulation steps below methodologically extend the theoretical model described in prior work on the simplicial model.

In this scenario, the goal of a CANCER SCREENING change agent in Los Angeles is to strategically diffuse a YouTube video about cancer screening and trigger recipients to click the 'Like' button and re-post the YouTube video on their Facebook pages (the request to click the 'Like' button and re-post is embedded in the YouTube video as part of the campaign message; a viewer can 'like' or re-post the video by clicking the 'Like' button right below the YouTube video or the 'Share' button and select to 'share' via Facebook) so local members of their bonded clusters will be exposed to the intervention message, and potentially do the same after hearing the request embedded in the video. 'Liking' and re-posting a YouTube video are common practices among Facebook users, and it shows support of and

commitment to the CANCER SCREENING message; similar to raising hands in a traditional STOP AIDS meeting and organizing more small group meetings). This intervention is an example of what Southwell (2013) calls an overt endorsement in his discussion of various information-sharing behaviors.

Burt's (1992) research on structural holes suggests that a network's connection to an external network through a single tie is necessary for information flow into a target network naturally. However, since bonded clusters can be computationally identified on Facebook, a change agent in our proposed model can initiate the intervention by going directly to target simplicial diffusers to get the intervention message to target groups, thus bypassing the need to have an existing tie. Although the simplicial model is used here to identify socially bonded clusters as the strategic targets of intervention, by virtue of posting a YouTube video on one's Facebook page, one's 'friends' outside of one's bonded clusters will be exposed to the intervention video as well, thus potentially spreading the video further due to weak ties. The intervention can subsequently spread through what Putnam (2000) calls bridging social capital. However, a change agent's target network is assumed to be locally based, when an intervention is by a community health center. The implementation of the proposed model is discussed in algorithmic steps below.

*Step one.* This step consists of the mining of network data from Facebook among specific target populations in the City of Los Angeles, so that one can build the simplicial complex; this complex is denoted by  $C_0$ , associated with the city. On Facebook, users can reveal their geographic locations and demographic information. In this case, one can filter users with the 'location' labeled as 'Los Angeles'. If a change agent would like to further narrow the target community based on gender, the change agent can also filter 'sex' as 'female' if the target is specifically for breast cancer screening and 'males' if the campaign is about colorectal cancer screening, testicular self-examination, etc. Note that the change agent does not have to restrict the target simplicial diffusers to one gender, because interest and willingness to promote a gender-specific cancer fortunately is not gender restricted. Males are equally likely to 'Like' or re-post a breast cancer screening YouTube video. We mention the ability to filter for multiple



demographic indicators only for illustration purposes. A researcher can also choose not to filter with any geographic location on Facebook, but the network data extracted may be too big for processing without the use of a supercomputer (Brynjolfsson & McAfee, 2014; Kee, Craddock, Blodgett, & Olwan, 2011). This step may have to be repeated over time as linkages are established, dissolved, and reconstituted. At this stage the change agent is not included in the complex that represents the city.

*Step two.* This step is the core of the algorithm. It has been established that every simplicial complex can be represented by a set of polynomials (in mathematical jargon, it is called an ideal of polynomials). In a simplified fashion, each 0-simplex  $A, B, C, D$ , etc. can be associated with a lower case letter, sometimes called a variable, such as  $a, b, c, d$ , etc. Then the 1-simplex geometrically represented by the segment  $[A,B]$  (i.e., a linkage between  $A$  and  $B$ ) is algebraically represented by the second degree polynomial  $ab$ , and so on. At the end of this process, the entire simplex is written as a set of polynomials, or more precisely, as an *ideal* of polynomials. In the program of research that builds on this study, our research team reveals that the ideal associated with a simplicial complex can be used to automatically seek the simplicial diffusers of a simplicial complex. Mathematically, this corresponds to the search for what is known as the primary decomposition of the ideal. The primary decomposition of the ideal offers several sets of nodes, all of which cover the entire complex, as explained in Proposition 7.

We will continue with a scenario of colorectal cancer screening among the male population in Los Angeles, an example of the decomposition, a case that illustrates a simplicial diffusing set. In accordance with the polynomial notation, the use of capital letters for nodes is dropped, and only lower case variables are used instead. We model a set of Facebook connections composed of Hector ( $h$ ) and William ( $w$ ), together with their respective friends Armando, Bryan and Cho, Denzel ( $a, b$ , and  $c, d$ ), whom they have met in their own personal lives. Suppose that their respective friends are not friends on Facebook although Armando and Bryan are aware of Cho and Denzel, and vice versa. However, they only communicate with their friends Hector and William on Facebook, such that the set of online connections can be represented by the schema at Time 1 on the left in Figure 2 in Appendix B. The three

social clusters in this complex are  $\{h,w\}$ ,  $\{a,b,h\}$ , and  $\{c,d,w\}$ .

What are the simplicial diffusing sets associated with this complex on Facebook? According to Proposition 7, a simplicial diffusing set  $T$  (i.e., a minimal cover) is defined as a set of nodes in the complex such that:

1. Each facet (or maximal face) in the complex has at least one element in  $T$ , and
2. No subset of  $T$  satisfies the previous property.

Looking for minimal covers in the complex at Time 1 in Figure 2, we can conclude that the groups  $\{h,w\}$ ,  $\{a,b,w\}$ , and  $\{c,d,h\}$  are potential simplicial diffusing sets for this complex on Facebook. Indeed, one can see that in all these three situations, conditions (1) and (2) are satisfied. The set  $\{h,w,a,b\}$  also satisfies condition (1), but it is not minimal in the sense that it contains  $\{h,w\}$ ; therefore, it does not satisfy condition (2). This result could also be computed quickly using the primary decomposition of the ideal  $(ha, hb, hw, wc, wd)$ .

The complex at Time 1 in Figure 2 can be interpreted by the analysis that if the CANCER SCREENING intervention is to be diffused within this set of nodes on Facebook, there are three possible scenarios. In the first scenario, corresponding to the set  $\{h,w\}$ , Hector and William receive the CANCER SCREENING YouTube video in a direct Facebook message from the change agent, they watch it, press the ‘Like’ button, re-post the YouTube video on their own Facebook pages, exposing their four friends to the same intervention message and trigger them to do the same. In the second scenario,  $h$ ’s friends  $a$  and  $b$ , as well as  $w$ , receive the Facebook message, watch the video, press the ‘Like’ button, re-post the video, and exposing the video to their friends  $h$ ,  $c$ , and  $d$ . It is easy for them to expose the video to  $h$  and potentially trigger him to do the same. Also,  $c$  and  $d$  will watch, ‘like,’ and re-post the video due to their friend  $w$ . Symmetrically, in the third scenario,  $c$  and  $d$  can become allied, so to speak, with  $h$  on Facebook (although without a direct link) because they all receive the Facebook message from CANCER SCREENING, watch the YouTube video, press the ‘Like’ button, and they re-post the video and trigger  $w$ ,  $a$ , and  $b$  to do the same.

If all six individuals are equally likely to further disseminate the intervention video, then the first scenario may be the most efficient because the diffusion effort is taken on by the smallest set of individuals (only two,  $h$  and  $w$ ). In addition, the global degree of information diffusion of both  $h$  and  $w$  on the set of Facebook connections is  $i(w)=i(h)=4$  (i.e.,  $1+1+1+1/2+1/2$ ; see Proposition 5), making a total value of 8 for the set  $\{w,h\}$ . The degree of information diffusion of, say,  $a$  (or equivalently, of any other less influential member of the set) is  $5/2$  (i.e.,  $1+1/2+1/2+1/4+1/4$ ), so that each of the alternative simplicial diffusing sets  $\{a,b,w\}$  and  $\{c,d,h\}$  has a total degree of information diffusion of 9 (i.e.,  $5/2+5/2+4$ ). One can choose to determine the potential effectiveness of each group by either the *average* or the *total* degree of information diffusion. If the average degree is the choice, then  $\{h,w\}$  is optimal, as the average value of their degree of information diffusion being 4 (i.e.,  $8/2$ ), as opposed to 3 (i.e.,  $9/3$ ) in the other two scenarios.

Now, we can compare this example with the case in which the social cluster of  $h$  is tightly connected. Suppose that his two friends become Facebook friends and the three of them often communicate and exchange information on Facebook among themselves, read and comment on each other's Facebook walls and status updates, join and participate in similar Facebook groups, 'Like' overlapping Facebook pages, indicate that they attend the same offline events advertised and organized via Facebook, tag each other in group pictures, videos, etc. (see earlier discussion on possible expressions of uniting conditions on Facebook, p. 5). Meanwhile,  $w$  and  $c$  get along and frequently communicate with each other on Facebook,  $c$  and  $d$  become Facebook friends, but  $d$  recently defriended  $w$  due to an argument. We represent this new situation at Time 2 for the set of Facebook connections  $\{a,b,c,d,h,w\}$  in the diagram on the right in Figure 2.

Observe that Hector's original set of Facebook connections is now modeled with a 2-simplex  $abh$ , a full triangle to represent a higher-dimensional group. The primary decomposition algorithm in this case returns a much longer list of possible simplicial diffusing sets for the complex on the right in Figure 2. Five of them have three members in it, while only one has two nodes:  $\{h,c\}$  (see Table 1). This

smallest set is a better candidate to be the simplicial diffusing set for this complex of six members on Facebook. Note that a simplicial diffusing set may *not* be a higher-dimensional group in itself. A computation with CoCoA software shows the average degree of information diffusion for these sets (see Table 1). Recall from Proposition 5 that the degree of information diffusion along a path of social distance  $d$  is  $(1/2)^{(d-1)}$ , and from Proposition 6 that an  $n$ -dimensional simplex has vertices whose social distance is  $1/n$ . To calculate the information diffusion from  $h$  to  $a$ , we add the degrees of information diffusion coming from both the direct link connecting  $h$  to  $a$ , which gives  $(1/2)^{(0.5-1)} = 1.4$  (path  $h-a$  belongs to a 2-simplex;  $d=1/2=0.5$ ), and the value obtained following (simplicial) double-link through  $b$ , which gives  $(1/2)^{(1-1)}=0$  ( $d=1$  because the distance of path  $h-b$ =distance of path  $b-a=1/2$ ). The same holds, symmetrically, for the total degree of information diffusion from  $h$  to  $b$ . Therefore, with step two, a list of possible simplicial diffusing sets within a simplicial complex on Facebook is identified via primary decomposition. They are ranked based on their average degree of information diffusion on the entire complex. *Step three* consists of restricting attention to those simplicial diffusing sets that are minimal, i.e., that contain the smallest possible number of nodes. Those will most likely produce a higher average degree of information diffusion on the entire complex. The change agent wants to identify small sets of nodes (as simplicial diffusers) on who will allow him to reach out to the entire simplicial complex on Facebook.

Assuming we have now identified  $N$  simplicial diffusing sets,  $T_1, \dots, T_N$ , each of which contains  $M$  nodes. In *step four*, one can build  $N$  new simplicial complexes. For each simplicial diffusing set  $T_i$  construct the complex  $C_i$  ( $i=1, \dots, N$ ) consisting of the original complex  $C_0$ , plus all the 1-simplices that represent the connections between the change agent and the nodes of the simplicial diffusing set  $T_i$ . Note that the complex  $C_i$  has exactly  $M$  new 1-simplices.

In *step five* we compute the global degree of information diffusion of the change agent on Facebook for each simplicial complex. It is natural to seek the complex in which the change agent has the largest degree of global information diffusion on Facebook. The simplicial diffusing set that has been

used to obtain that complex is the output of the algorithm, and it tells the change agent which members in the Facebook community he needs to reach out to (or partner with), to create a solid coalition, which will promote health intervention in the target community on Facebook. It could be the case also that there is more than one set which maximizes the total degree of information diffusion of the change agent, in which case the algorithm discussed returns all the simplicial diffusing sets that generated the maximum global degree of information diffusion.

As previously explained, the simplicial approach taps into the potential of strong ties to trigger users to click the ‘Like’ button and re-post the YouTube video, mobilizing users to promote a health intervention campaign, in a fashion similar to the historical protest described by Gladwell (2010). In the meantime, the traditional approach by Valente and colleagues (1995; 1996; 2007) takes advantages of opinion leaders and their weak ties to spread health messages to a large number of members within a social system, in a way similar to the arguments advanced by Granovetter (1973;1983) and Burt (1992). We demonstrate in the example simulation below to show how both approaches are complementary to each other, and one may yield a better result under certain circumstances.

### **An Example Computational Simulation for Strategic Diffusion**

We offer an example of how this algorithm would work in practice. To make the point without unnecessary complexity, we limit the total number of users in the following scenario to 23 members in a hypothetical Los Angeles male population for a colorectal cancer screening campaign, a specific effort of CANCER SCREENING. The example below uses a randomly generated system with three social aggregations (i.e., simplicial complexes) of Facebook connections with 5, 8, and 10 members, respectively.

To demonstrate the difference between opinion leaders and simplicial diffusers, in this example it is assumed that in each social aggregation there is an additional digital opinion leader who is connected to everyone within his social aggregations. Valente (1995) describes an opinion leader as someone with the largest number of connections within a group. Therefore, this digital opinion leader is also the individual

who holds each group together, as some nodes in each group only connect to him.

Carrington and Scott (2011) state, “Networks containing more than 20 or so points are difficult to draw accurately and legibly as sociograms” (p. 5). Pictorially, we simplified an instance of such a situation with the diagram in Figures 4 and 5 in Appendix D. The three big boxes represent the three different social aggregations (i.e., simplicial complexes). Although the article treats the visual representation in Figure 3 as three simplicial complexes, one can also argue that the entire network make up a larger simplicial complex.

As a shortcut, the fact that each digital opinion leader  $D_i$  is connected to every member of his social aggregation is pictorially represented by a single edge from the  $D_i$  node to the box enclosing members within the social aggregation. The two seemingly isolated nodes in the first box on the left are actually connected to the node  $D_1$ . The boxes themselves do not constitute a part of the simplicial complex, but simply represent all the 1-dimensional connections between the digital opinion leader and each of the members of his social aggregation. The computer simulation was allowed to determine those connections in a random fashion. For the sake of simplicity, the dimension of such simplices was limited to triangles (as a higher-dimensional group), although this is not a requirement for the algorithm.

The goal of the CANCER SCREENING change agent is to reach out to strategic individuals (simplicial diffusers and opinion leaders) with a YouTube video that promotes colorectal cancer screening, encourage the recipients to ‘Like’ the video, and request them to re-post the video on the Facebook pages in order to further expose the message to other males on Facebook. When the simplicial algorithm is run for the system in Figure 3, several possible solutions are obtained.

First of all, there are a total of 192 different simplicial diffusing sets on Facebook to which the CANCER SCREENING change agent can reach out to. The number of people in each set ranges from 6 to 17, but only 16 of such simplicial diffusing sets are actually formed by 6 members (the minimal size possible). According to Proposition 7, we will focus on these 16 smallest simplicial diffusing sets. Then we also added the three digital opinion leaders  $D_1$ ,  $D_2$ , and  $D_3$  into each set because they have the highest

degree centrality. The 16 sets are then labeled as  $T_1, \dots, T_{16}$ . Each  $T_i$  includes the six cluster diffusers and the three digital opinion leaders in Figure 3, making each set consisting of nine individuals.

By applying Proposition 6, we show in Table 2 the global degree of information diffusion and average degree of information diffusion (i.e., global degree of information diffusion divided by 9 members in each simplicial diffusing set) that a change agent would gain by joining the set  $T_i$ . As we can see from Table 2, the global degree of information diffusion of the CANCER SCREENING change agent can be optimized by joining the simplicial diffusing set  $T_7$  or  $T_{11}$ .

Based on the example simulation, Figure 4 shows that a change agent (let's call him Dean) can identify the socially bonded clusters within the target social aggregations and strategically select the smallest set of simplicial diffusers (with strong ties; those who have circles around them) and opinion leaders (with weak ties;  $D_1$ ,  $D_2$ , and  $D_3$ ) to accelerate the diffusion of a colorectal cancer screening video within a hypothetical male population in Los Angeles on Facebook. We include both opinion leaders and simplicial diffusers in this example simulation to demonstrate that the simplicial method of mathematically and computationally identify simplicial diffusing sets is complementary to the traditional approach of partnering with opinion leaders within a target community. In other words, both approaches can be applied simultaneously to further optimize health information diffusion.

At this point, one may wonder how the simplicial approach compared to the traditional approach of targeting simply opinion leaders. Are there specific circumstances that make one approach more powerful than the other? To compare and contrast the two approaches, let us confine our attention to the simplicial complex connected to  $D_2$  (i.e., the middle box) in Figure 3. We re-drew such a box in Figure 5 in Appendix F, showing all the pairwise connections and higher-dimensional groups in this simplicial complex. We will consider the simplicial complex whose nodes are  $D_2$ ,  $x$ ,  $y$ , and  $\alpha, \beta, \gamma, \delta, \varepsilon, \theta$ . The reason we use Latin letters 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 traditional point of view,  $D_2$  appears to be the opinion leader because it has the highest degree centrality. We therefore want to calculate the degree of information diffusion of  $D_2$  and compare it with the degree of information diffusion of the pair  $\{\alpha, \beta\}$  (i.e., mathematically a minimal cover). In order to follow the calculations below, we will assign degrees of information diffusion of 1 to any direct link (such as  $D_2-x$ ), 0.5 to any double-link (such as  $D_2-\theta-\alpha$ ), 1.4 (to approximate  $\sqrt{2}$ ) to any simplicial link (such as  $\alpha-\beta$ ), 0.7 for simplicial double-link (such as  $\alpha-\beta-\gamma$ ), and 0.6 to what we call a mixed double-link (such as  $D_2-\alpha-\beta$ ). The existence of a two-dimensional simplex (as in the triangle connecting  $\alpha$ ,  $\beta$ , and  $\gamma$ ) shortens, so to speak, the distances between vertices, and accordingly increases the degrees of information diffusion.

Recall from Proposition 5 that the degree of information diffusion along a path of social distance  $d$  is  $(1/2)^{(d-1)}$ , and from Proposition 6 that an  $n$ -dimensional simplex has vertices whose social distance is  $1/n$ . The degree of information diffusion along a double-link is 0.5. However, along a simplicial double-link (such as  $\alpha-\beta-\gamma$ ), the distance from  $\alpha$  to  $\beta$  is 0.5 because they belong to a full triangle (see Proposition 6). The degree of information diffusion of  $\alpha$  on  $\beta$  is then  $(1/2)^{(0.5-1)} = (1/2)^{(-0.50)} = 1.4$ . Furthermore, the path  $\alpha-\beta-\gamma$  is a simplicial double-link. Therefore, the degree of information diffusion that  $\alpha$  has on  $\gamma$  is 0.7, half of the direct flow from  $\alpha$  to  $\beta$ . Since the mixed-second order links are in between a regular double-link and a simplicial double-link, we approximate the value to be the average value between the two, which is 0.6.

Furthermore, in keeping with a reasonable assumption by Christakis and Fowler's (2009) cited for Proposition 2, we 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  $\delta$  and then  $\beta$ , but we will not consider further information diffusion that can be achieved by going from  $D_2$  to  $\delta$ , to  $\beta$ , to  $\varepsilon$ .



With these two caveats in mind, the degrees of information diffusion are 12.8 for  $\alpha$ , 15.9 for  $\beta$ , 16.2 for  $D_2$ , and 28.7 for  $\{\alpha, \beta\}$ . One could however argue that the average degree of information diffusion for  $\{\alpha, \beta\}$  is less than  $D_2$ . Whether this is an important objection is debatable, because the set  $\{\alpha, \beta\}$  still offers much better degree of information diffusion than  $D_2$  alone.

Furthermore, if we remove the vertices  $x$  and  $y$ , the degrees of information diffusion become 11.8 for  $\alpha$ , 14.9 for  $\beta$ , only 14.2 for  $D_2$ , and 28.7 for  $\{\alpha, \beta\}$ . Thus, despite the fact that  $D_2$  remains the node with the largest degree centrality, we have now found that  $\beta$  alone has a higher degree of information diffusion than  $D_2$ . Note also that the values for  $D_2$  dropped by 2 (the number of terminal nodes removed), but 1 for  $\alpha$  and  $\beta$  each, making  $\beta$  increasingly more powerful than  $D_2$  in a situation where there are more simplexes than terminal vertices.

Another comment that may be worthwhile is that the difference between the two methods becomes larger and more visible as we increase the number of higher-dimensional simplexes. Indeed, it is exactly when vertices belong to such higher-dimensional simplexes that the calculation will show increased degree of information diffusion. So, for example, if  $\gamma$  and  $\varepsilon$  were also connected, and  $\beta, \gamma, \varepsilon$  make up the third full triangle, the degree of information diffusion of  $\beta$  would grow again more than the one of  $D_2$ , thus making the simplicial method even more powerful. In this case, indeed, the degree of information diffusion of  $D_2$  would not change, while the degree of information diffusion of  $\beta$  would grow by another  $\sqrt{2}$  or 1.4, so that even in the case of the simplex with only one terminal vertex the average degree of information diffusion of  $\{\alpha, \beta\}$  would end up being higher than the degree of information diffusion of  $D_2$  alone.

Finally, it is important to note that not all individuals who receive the CANCER SCREENING YouTube video will press the 'Like' button and re-post it on their own Facebook page and expose others within their bonded clusters to the intervention. Therefore, it would be prudent for the change agent to build in some redundancy in the strategic selection of partners. That is why we included both simplicial diffusers and opinion leaders in our modeling. Moreover, a change agent can reach out to both  $T_7$  and  $T_{11}$

simultaneously. If a change agent would like to further include the ‘next best’ set besides  $T_7$  and  $T_{11}$ , s/he can also include  $T_3$ , as it is the set with the second highest degrees of average information diffusion according to Table 2. We discuss key advantages of the simplicial approach in the implications below.

### **Conclusion and Future Directions**

We are interested in *how the participatory nature of Facebook can be leveraged to improve cancer prevention and control*. More specifically, we posed the research question: *How can socially bonded clusters be mathematically modeled on Facebook to strategically accelerate peer-to-peer information diffusion for community health interventions?* We proposed an alternative approach to accelerate information diffusion through simplicial diffusing sets and their corresponding bonded groups within a larger Facebook community. Next, we conclude with a few implications and limitations of our model.

First, this article builds on the simplicial model of social aggregation (Kee, et al., 2013; Mannucci, et al., 2006; Sparks, et al., 2014) and introduces the constructs of *simplicial diffusers* and *simplicial diffusing sets* in a target Facebook community. These constructs are further defined mathematically as *nodes in simplexes* and *minimal covers* embedded within a simplicial complex. These concepts extend classical network analysis to consider higher-dimensional groups and communities.

Second, this article provides a mathematical approach for calculating the degree of information diffusion as a criterion for selecting simplicial diffusing sets to aid in diffusing health intervention messages on Facebook. These equations build on sound assumptions and the foundations of classical network literature, but extending them to higher-dimensional simplexes and simplicial complexes.

Third, the use of simplicial diffusing sets to promote a health intervention takes advantage of the frequent internal communication patterns within socially bonded clusters. 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 clusters to attend and respond to the intervention message. Also, health communication research has shown that the family

plays an important role in processing complex health issues (Sparks, 2007), and our approach allows for that.

Fourth, individual opinion leaders 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 opinion leaders and share it among a group of simplicial diffusers. The results could be more rapid diffusion. In a community with 100 members, a difference of 10 more simplicial diffusers instead of simply 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.

Fifth, Facebook or a similar social media platform can more easily be used for filtering users based on demographic and social indicators, such as 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).

The proposed model is not without limitations. The simplicial model recommends simplicial diffusers and simplicial diffusing sets based on members' structural positions. Another structural criterion that can be incorporated in future research is to distinguish between primary and secondary simplicial diffusing sets based on whether members have internal connections. While the current conceptualization of simplicial diffusing sets is that simplicial diffusers do not need to make up a higher-dimensional group, having internal connections intuitively makes sense to further enhance a diffusion effort.

Moreover, we did not take into account the individual characteristics of simplicial diffusers, such as their trust with their cluster members. Similarly, with emerging online influence ranking indices, such as Klout, PeerIndex, Twitter Grader (Kadushin, 2012), the selection of simplicial diffusers can be further enhanced with these online influence ranking indices. On another note, future research should also look into the impact of signal-to-noise ratio on any single post. If a user has the tendency to post a lot, his/her

posts may receive a different degree of peer attention compared to others who post only occasionally.

Third, Proposition 4 and Proposition 5 allow for the addition of direct and indirect information diffusion into a number that is not capped. A valid concern is that it may lead to some cases where someone has an unrealistic degree of information diffusion over another person simply due to having many mutual friends. Future research should empirically test how the addition of information diffusion should be capped, in addition to the possible solution proposed by Christakis and Fowler's (2009).

Fourth, Southwell (2013) argues that individual level, community-level, and content-level factors collectively influence the ways people share and diffuse health and science related information within their social networks. He notes that an individual's need for sensation seeking can positively impact one's likelihood to share health and science information; negatively framed messages seem to more likely travel in more interconnected groups. This study only looks at how socially bonded groups play a role in information diffusion at the community-level. Future research should look into how individual- and content-level factors influence information diffusion within socially bonded groups.

Fifth, the accessibility of Facebook data for mining and analysis remain less open compared to other social media platforms such as Twitter. While most Facebook users keep their information (including their Friendship connections) private, some demographics (e.g., teens) include many users (40%) who do share this information in a way that the data is accessible via the Facebook API. More importantly, the simplicial approach could potentially be implemented by Facebook themselves, without having to compromise any private information. In other words, Facebook could develop a tool that would allow researchers and change agents to generate a list of simplicial diffusers and allow researchers and change agents to reach out to those users. In fact, Facebook is hosting a private sociology conference in August 2014 in the Bay Area, CA (Hernandez, 2014). The company will be showcasing its tools and software stack to social scientists they recruit to use its data for social research, and 'user engagement' is identified as of key interest to Facebook (Zhou, 2014). This development raises interesting possibilities, including the use of the simplicial approach in targeted advertising campaigns both in and outside of the

health domain.

Finally, Bailey, Leonardi, and Barley (2011) warn against the ‘lure of the virtual’ and suggest treating a simulation as simply a simulation, and not a picture of full reality. We take this caution seriously and note that the simplicial simulation presented in this article is merely an analytical attempt to extend current theoretical understanding of networks, although we believe that it also has important implications for practical knowledge of innovation diffusion strategies.

STOP AIDS was extremely successful at curtailing the spread of HIV/AIDS back in the 1980s. Rogers (2004) reported that about 7,000 gay men were recruited and trained in the STOP AIDS meetings in small groups and an additional 30,000 gay men were estimated to have been reached by this innovation diffusion strategy. Reported HIV infections decreased from 8,000 new cases in 1983 to 650 in 1985. Reported unprotected sex dropped a significant 44 percent between 1983 and 1987, from 71 percent to 27 percent within four years. In proposing a computational simulation approach, we would like to echo Dearing’s (2004, p. 32) argument below in the same spirit of HIV for cancer prevention and control:

In a computer simulation, Stover et al. (2002), publishing in *The Lancet*, found that if proven programs in HIV/AIDS prevention were actually adopted, 29 million new infections could be prevented over the next 8 years. If we can affect even a small proportion of the practitioner decisions that will be made about health programming options in coming years, then our work will be well worth the effort.

This article presented a mathematical and computational approach to model and recommend cancer prevention, control, and intervention strategies involving identifying socially bonded clusters and simplicial diffusers within a Facebook community. The simplicial model of social aggregation and the mathematical theory of simplicial complexes are applied to allow for a precise and flexible mathematical description of key higher-dimensional constructs, including higher-dimensional groups, social aggregations, and simplicial diffusing sets. It is our hope that this article sparks future research in the development and modeling of higher-order dimensions of social aggregations as an important characteristic in a more comprehensive understanding of communication networks for innovation diffusion and community health intervention.

## References

- Arabie, P., & Carroll, J. D. (1989). Conceptions of overlap in social structure. In J. Freese, D. R. White & A. K. Romney (Eds.), *Research methods in social network analysis* (pp. 367-392). Fairfax, VA: George Mason University Press.
- Bailey, D. E., Leonardi, P. M., & Barley, S. R. (2011). The lure of the virtual. *Organization Science*, 1-20.
- Beck, J. (1979). *Simplicial sets and the foundations of analysis*. Paper presented at the Applications of sheaves (Proc. Res. Sympos. Appl. Sheaf Theory to Logic, Algebra and Anal., ), University of Durham, Durham, 1977.
- Bertrand, J. T. (2004). Diffusion of innovations and HIV/AIDS. *Journal of Health Communication*, 9, 113-121.
- Borgatti, S. P., Everett, M. G., & Shirey, P. R. (1990). LS sets, lambda sets and other cohesive subsets. *Social Networks*, 12(4), 337-357.
- Borgatti, S. P., & Lopez-Kidwell, V. (2011). Network theory. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 40-54). Long Angeles: Sage.
- Boyd, J. P. (1991). *Social semigroups: A unified theory of scaling and blockmodelling as applied to social networks*: George Mason University Press Fairfax, VA.
- Breiger, R. L., Boorman, S. A., & Arabie, P. (1975). An algorithm for clustering relational data with applications to social network analysis and comparison with multidimensional scaling. *Journal of Mathematical Psychology*, 12(3), 328-383.
- Brooke-Taylor, A., & Testa, D. (2013). *The infinite random simplicial complex*. Heilbronn Institute for Mathematical Research. University of Bristol and Kobe University. Retrieved from <http://arxiv.org/pdf/1308.5517.pdf>
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*: WW Norton & Company.
- Burt, R. S. (1992). *Structural holes: The social structure of competition*. Cambridge, MA: Harvard University Press.
- Carrington, P. J. (2011). Crime and social network analysis. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 236-255). Long Angeles: Sage.
- Carrington, P. J., & Scott, J. (2011). Introduction. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 1-8). Long Angeles: Sage.
- Carroll, J. D., & Arabie, P. (1998). Multidimensional scaling. In M. H. Birnbaum (Ed.), *Measurement, judgment, and decision making* (pp. 179-250). San Diego, CA: Academic Press.
- Chou, W.-Y. S., Hunt, Y. M., Beckjord, E. B., Moser, R. P., & Hesse, B. W. (2009). Social media use in the United States: Implications for health communication. *Journal of Medical Internet Research*, 11(4).
- Chou, W.-y. S., Prestin, A., Lyons, C., & Wen, K.-y. (2013). Web 2.0 for health promotion: reviewing the current evidence. *American Journal of Public Health*, 103(1), e9-e18.
- Christakis, N. A., & Fowler, J. H. (2009). *Connected: The surprising power of our social networks and how they shape our lives*: Hachette Digital, Inc.
- Chua, V., Madej, J., & Wellman, B. (2011). Personal communities: The world according to me. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 101-115). Long Angeles: Sage.
- Daly, J. A. (2011). *Advocacy: Championing ideas and influencing others*. New Haven, CT: Yale University Press.
- Dearing, J. W. (2004). Improving the state of health programming by using diffusion theory. *Journal of Health Communication*, 9, 21-36.

- 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.
- Dekker, A. (2001). *Visualisation of social networks using CAVALIER*. Paper presented at the Proceedings of the 2001 Asia-Pacific symposium on Information visualisation.
- Diani, M. (2011). Social movements and collective action. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 223-235). Long Angeles: Sage.
- Everett, M. G., & Borgatti, S. P. (1998). Analyzing clique overlap. *Connections*, 21(1), 49-61.
- Eysenbach, G. (2008). Medicine 2.0: Social networking, collaboration, participation, apomediation, and openness. *Journal of Medical Internet Research*, 10(3), e22.
- Faridi, S. (2002). The facet ideal of a simplicial complex. *manuscripta mathematica*, 109(2), 159-174.
- Freeman, L. C. (2011). The development of social network analysis -- with an emphasis on recent events. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 26-39). Long Angeles: Sage.
- Freeman, L. C., Romney, A. K., & Freeman, S. C. (1987). Cognitive structure and informant accuracy. *American Anthropologist*, 89(2), 310-325.
- Gladwell, M. (2010, October 4). Small change: Why the revolution will not be tweeted. *The New Yorker*, 4, 42-49.
- Gould, J., & Fernandez, J. (1989). Structures of mediation: A formal approach to brokerage in transaction networks. *Sociological Methodology*, 89-126.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 1360-1380.
- Granovetter, M. S. (1983). The strength of weak ties: A network theory revisited. *Sociological Theory*, 1(1), 201-233.
- Hawn, C. (2009). Take two aspirin and tweet me in the morning: How Twitter, Facebook, and other social media are reshaping health care. *Health Affairs*, 28(2), 361-368.
- Hernandez, B. E. (2014). Facebook's private sociology conference. *NBC News Bay Area*. Retrieved from <http://www.nbcbayarea.com/contact-us/>
- Idriss, S. Z., Kvedar, J. C., & Watson, A. J. (2009). The role of online support communities: Benefits of expanded social networks to patients with psoriasis. *Archives of Dermatology*, 145(1), 46-51.
- Kadushin, C. (2012). *Understanding social networks: Theories, concepts, and findings*. New York: Oxford University Press.
- Katz, E., & Lazarsfeld, P. F. (1955). *Personal Influence: The part played by people in the flow of mass communications*. New Brunswick, NJ: Transaction Publishers.
- Kee, K. F., Craddock, L., Blodgett, B., & Olwan, R. (2011). Cyberinfrastructure inside out: Definitions and influences shaping its emergence, development, and implementation In D. Araya, Y. Breindl & T. Houghton (Eds.), *Nexus: New intersections in Internet research*. (pp. 157-189). New York: Peter Lang.
- 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(1), 35-58. doi: 10.1080/01463373.2012.719566
- Kelly, J. A., St Lawrence, J. S., Diaz, Y. E., Stevenson, L. Y., Hauth, A. C., Brasfield, T. L., . . . Andrew, M. (1991). HIV risk behavior reduction following intervention with key opinion leaders of population: an experimental analysis. *American Journal of Public Health*, 81(2), 168-171.
- 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(3), 315-350.
- Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., . . . Gutmann, M. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915), 721-723.

- Lewin, K. (1947). Frontiers in group dynamics II: Channels of group life; social planning and action research. *Human Relations*, 1(2), 143-153.
- Mannucci, M. A., Sparks, L., & Struppa, D. C. (2006). Simplicial models of social aggregation I. Retrieved from <http://arxiv.org/abs/cs/0604090>.
- Marin, A., & Wellman, B. (2011). Social network analysis: An introduction. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 11-25). Long Angeles: Sage.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that Will Transform how We Live, Work, and Think*: Eamon Dolan/Houghton Mifflin Harcourt.
- Mische, A. (2011). Relational sociology, culture, and agency. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 80-98). Long Angeles: Sage.
- Monge, P. R., & Contractor, N. S. (2003). *Theories of communication networks*. New York: Oxford University Press
- Munkres, J. R. (1984). *Elements of algebraic topology*. Menlo Park, CA: Addison-Wesley Publishing Company.
- Norman, C. D., McIntosh, S., Selby, P., & Eysenbach, G. (2008). Web-assisted tobacco interventions: Empowering change in the global fight for the public's (e) Health. *Journal of Medical Internet Research*, 10(5).
- Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043), 814-818.
- Provencher, S. W. (1976). An eigenfunction expansion method for the analysis of exponential decay curves. *The Journal of Chemical Physics*, 64, 2772-2778.
- Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. New York: Simon and Schuster.
- Ren, W., Zhao, Q., Ramanathan, R., Gao, J., Swami, A., Bar-Noy, A., . . . Basu, P. (2011). *Broadcasting in multi-radio multi-channel wireless networks using simplicial complexes*. Paper presented at the Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on.
- Rice, R. E. (2001). Diffusion of innovations and communication. In J. Schement (Ed.), *Encyclopedia of communication and information* (pp. 248-253). New York, NY: Macmillan Reference.
- Rogers, E. M. (2003). *Diffusion of innovations* (5th ed.). New York: Free Press.
- Rogers, E. M. (2004). A prospective and retrospective look at the diffusion model. *Journal of Health Communication*, 9(S1), 13-19.
- Sabidussi, G. (1966). The centrality index of a graph. *Psychometrika*, 31(4), 581-603.
- Singhal, A., & Rogers, E. M. (2003). *Combating AIDS: Communication strategies in action*. New Delhi, India: Sage.
- Southwell, B. G. (2013). *Social networks and popular understanding of science and health: Sharing disparities*: Johns Hopkins University Press.
- Sparks, L. (2007). Cancer care and the aging patient: Complexities of age-related communication barriers. . In H. D. O'Hair, G. L. Kreps & L. Sparks (Eds.), *Handbook of communication and cancer care*. Cresskill, NJ: Hampton Press.
- 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.
- Thackeray, R., Neiger, B. L., Hanson, C. L., & McKenzie, J. F. (2008). Enhancing promotional strategies within social marketing programs: Use of Web 2.0 social media. *Health Promotion Practice*, 9(4), 338-343.
- Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes*, 103(1), 37-52.
- Valente, T. W. (1995). *Network models of the diffusion of innovations*. Cresskill, NJ: Hampton.



- Valente, T. W., & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders. In P. Lopes & M. Durfee (Eds.), *The social diffusion of ideas and things* (pp. 55-67). Thousand Oaks, CA: Sage Publications.
- Valente, T. W., & Pumpuang, P. (2007). Identifying opinion leaders to promote behavior change. *Health Education & Behavior, 34*(6), 881-896.
- van der Hulst, R. C. (2011). Terrorist networks: The threat of connectivity. In J. Scott & P. J. Carrington (Eds.), *The SAGE Handbook of Social Network Analysis* (pp. 256-270). Long Angeles: Sage.
- Vance, K., Howe, W., & Dellavalle, R. P. (2009). Social internet sites as a source of public health information. *Dermatologic Clinics, 27*(2), 133-136.
- Wangberg, S. C., Andreassen, H. K., Prokosch, H.-U., Santana, S. M. V., Sørensen, T., & Chronaki, C. E. (2008). Relations between Internet use, socio-economic status (SES), social support and subjective health. *Health promotion international, 23*(1), 70-77.
- Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.
- Weimann, G. (1994). *The influentials: People who influence people*: SUNY Press.
- Wohlfeiler, D. (1998). Community organizing and community building among gay and bisexual men: The STOP AIDS project. In M. Minkler (Ed.), *Community organizing and community building for health*. New Brunswick, NJ: Rutgers University Press.
- Woolf, S. H. (2008). The power of prevention and what it requires. *JAMA: The Journal of the American Medical Association, 299*(20), 2437-2439.
- Zhou, L. (2014). EXCLUSIVE: Facebook plans event to recruit sociologists. *VB (VentureBeat) News*. Retrieved from <http://venturebeat.com/2014/06/07/exclusive-to-sell-ads-in-the-developing-world-facebook-is-hiring-sociologists/>
- Zubcsek, P. P., Chowdhury, I., & Katona, Z. (2014). Information communities: The network structure of communication. *Social Networks, 38*, 50-62. doi: <http://dx.doi.org/10.1016/j.socnet.2014.02.001>

## Appendix A

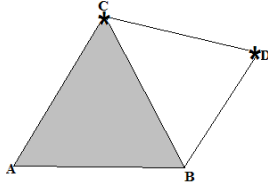


Figure 1: The full triangle  $[A,B,C]$  is a tightly bound or higher-dimensional group while the regular triangle  $[B,C,D]$  is simply a group based on pair-wise connections. The set of two nodes  $\{C,D\}$  is one of several possible covers for this complex consisted of A, B, C, and D.

## Appendix B

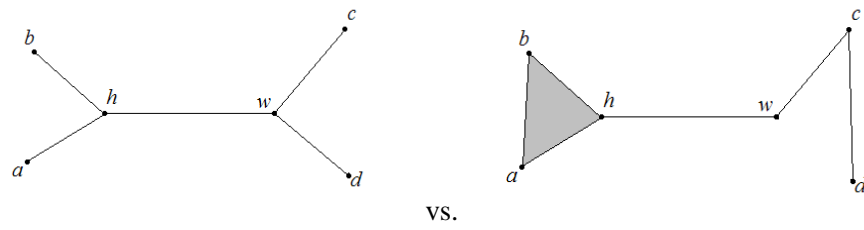


Figure 2: A social aggregation (i.e., a simplicial complex) on Facebook on the left at Time 1 where Hector ( $h$ ) and William ( $w$ ) each have two other friends Armando ( $a$ ) and Bryan ( $b$ ) as well as Cho ( $c$ ) and Denzel ( $d$ ), versus an alternative configuration of the social aggregation on the right at Time 2.

Appendix D

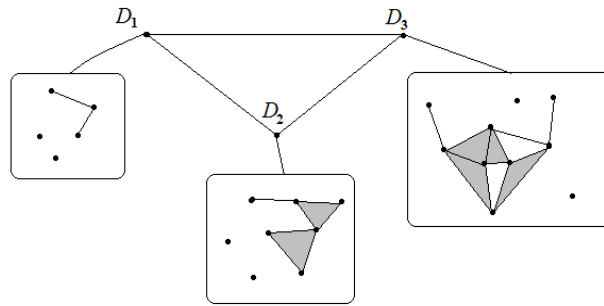


Figure 3: A simplified hypothetical Los Angeles male population on Facebook with three social aggregations and three digital opinion leaders  $D_1$ ,  $D_2$ , and  $D_3$

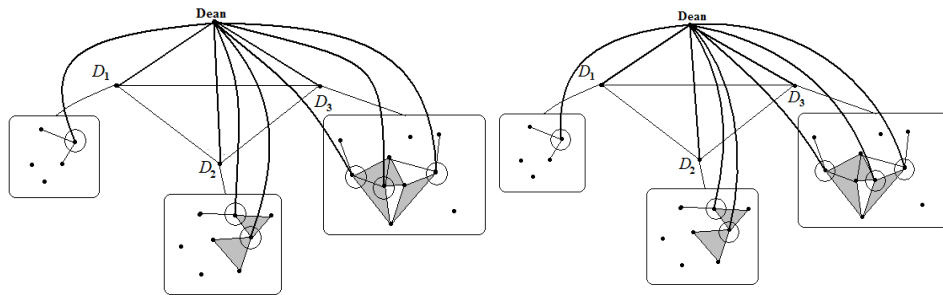


Figure 4: The two strategic sets of simplicial diffusers for the change agent, corresponding to  $T_7$  and  $T_{11}$

## Appendix F

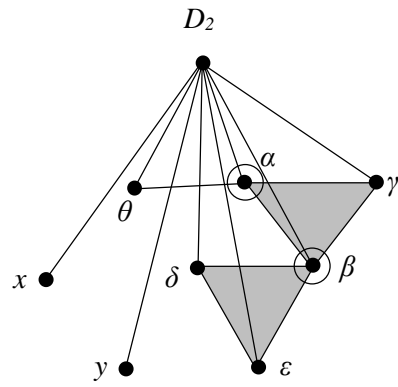


Figure 5: The simplicial complex of  $D_2$  with its pairwise connections and higher-dimensional groups.