Spillover within social groups: Understanding the contagions of messaging, habits, and other interventions within peer-groups

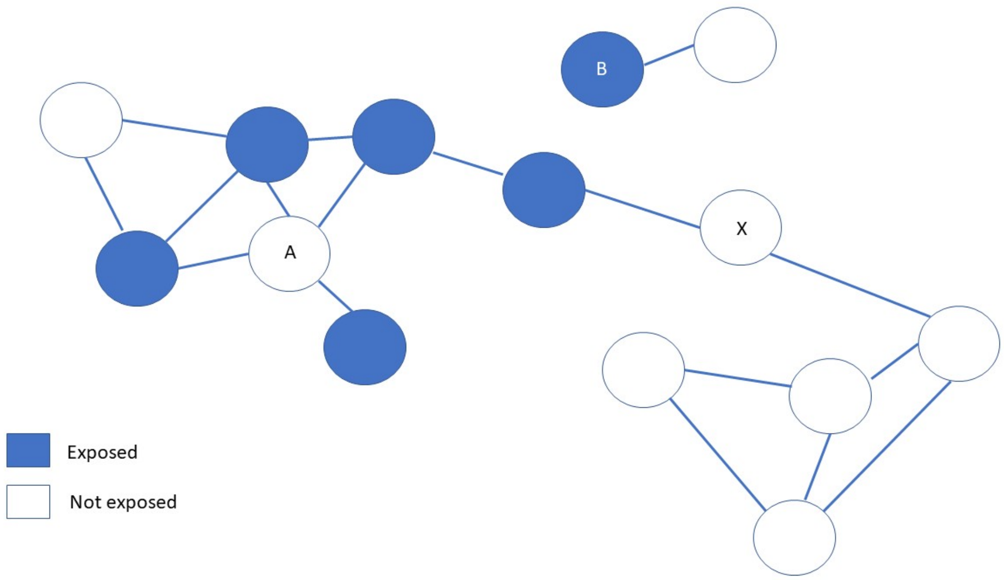
By Erica EM Moodie (McGill University) and Nema Dean (University of Glasgow)

We often hear the terms "social contagion" or "emotional contagion", for instance in the context of the now infamous experiment Facebook experiment[1] in which users' News Feeds were manipulated to determine the effect on individuals' moods. And yet many of the methods we have to understand the factors that affect our health – whether they might increase the risk of negative outcomes or in fact be protective – are assumed to affect only the individual who is directly exposed to those factors. Take, for example, two friends, Ahmed and Brenda. Statistically speaking, we typically assume that there is no *interference* between Ahmed's exposure and Brenda's outcome: Ahmed's outcome may be influenced by *his* exposure and Brenda's by *hers*, but neither is affected by the exposure of the other. By exposure, we could be talking about an intervention of some kind, e.g. prescribing a treatment or not, or we could be talking about something more negative like exposure to contracting a disease of some kind. A canonical example of when this can't be assumed is in the case of vaccines for contagious diseases: if all of Brenda's peers (including Ahmed) are vaccinated, her chance of contracting the disease are greatly reduced since she is surrounded by people who cannot catch it – and thus cannot transmit it to her.

Social networks are a summary of the patterns of connections existing between people (sometimes called actors or organisations in the statistical literature). These networks, or graphs, record important information about dependencies or associations between the people of interest. This can be as simple as recording pairs of our individuals who are connected to each other, e.g. a list of friendships or people following each other on social media. The statistical study of these types of dependency structure is often known as social (or statistical) network analysis[2,3].

As John Donne said, "no man is an island". Human beings exist in a constant state of connections of various kinds. We know that people are connected so the idea of assuming non-connectivity to try to understand the effect of an intervention seems quite nonsensical. In fact, ignoring this information when trying to understand the consequences of an action that can spread between connections could greatly bias the results found. Consider the figure below: if we compare the outcome of person A who did not receive an exposure and person B who did, explaining the effect of the exposure as the difference between the outcomes of the two will only make sense if person A is truly unexposed, and has not received some part of the exposure through spillover or "contamination" from his exposed connected partners: we could be underestimating the effect because we are not really comparing exposed and non-exposed but rather treated and partially treated. (Of course, individuals A and B must otherwise be comparable, so that differences in outcome are not attributable to differences in factors other than the exposure of interest)

[1] Kramer, Guillory, and Hancock (2014) Experimental evidence of massive-scale emotional contagion through social networks, *Proceedings of the National Academy of Sciences* **111**:8788-8790
[2] Kolaczyk (2009) Statistical Analysis of Network Data: Methods and Models. Springer
[3] Luke (2016) A User's Guide to Network Analysis in R. Springer



In cases where we are interested in studying the spread of disease, e.g. HIV transmission between sexual partners[4] or Hepatitis C virus between needle-sharing partners, we can use the network information to try to predict the pattern of disease spread through the network. Information about networks and their individuals can then be useful in targeting interventions at the points where they can be most beneficial. You might think that it would be obvious to target the people with the most connections to stop infections and this is one network attribute that can be useful to identify targets for intervention. However, we might also be interested in individuals who, while not having many connections, have important types of connections, e.g. a person like X in our figure has two sexual partners but if we stop infection with X, we can prevent it from spreading to an entire portion of the population (the four individuals in the lower right corner) whose only connection to the others is through X. Measuring different types of importance of connections or individuals in a network is part of what is offered by network analysis.

Given network information can be so important in the study of disease spread, surely analysis of interventions of other kinds that can affect the network or be affected by the network need to use this information too!

There are in fact many interventions that may be transmitted or diffused through social networks, a phenomenon that has been recognized by adolescent researchers who have leveraged the power of social bonds among teenagers to disseminate information through these networks through peer-led interventions. Examples include the ASSIST (A Stop Smoking In Schools Trial)[5] and STASH[6] (Sexually Transmitted infections And Sexual Health) studies, both of which used models in which a small number of

---

[4] Krivitsky & Morris (2017) Inference for social network models from egocentrically sampled data, with application to understanding persistent racial disparities in HIV prevalence in the US. *Annals of Applied Statistics*, **11**:427-455

[5] Hollingworth et al. (2012) Reducing smoking in adolescents: cost-effectiveness results from the cluster randomized ASSIST (A Stop Smoking In Schools Trial). *Nicotine & Tobacco Research* **14**:161-168

[6] Mitchell et al. (2020) STI prevention and sexual health in secondary schools: An exploratory study of a peer-led intervention (STASH). *Public Health Research* (in press)

students within schools were given targeted educational training and asked to spread the message amongst their peers. The "exposure" of interest in both cases was the targeted training, however the hope was indeed that this would spill over onto their peers, so that all students in the schools would see enhanced awareness and improved outcomes, even those who did not participate in the training. From a statistical view, we could simply try to understand whether, at a group or school level, the training of some students was effective, viewing each school as a unit of analysis. However, we might also or alternatively want a more fine-grained understanding of the effect of the exposure: how much did it affect those who were the direct recipients (the students who received the training)? What was the impact of having one or perhaps several friends receive the training for a student who did not? The first of these questions asks about the *direct effect* of the educational intervention while the second targets the *indirect effect*. How many students should receive targeted training to achieve some desired group level of the outcome? This final question requires a consideration of the *total effect*, which accounts for both the direct and indirect effects.

Similar questions arise in the context of a pandemic such as the one through which we are now living. To what extent do the hand-washing habits of one person in a family extend to another? Can targeted mental health support to care workers in nursing homes spill over to their families, or to the residents for whom they care, thereby easing the emotional effects of social/physical distancing? Further down the line, as a vaccine for the new coronavirus (SARS-CoV-2) is developed, understanding what proportion of a given social network is needed to confer broad group immunity may help to develop targeted vaccination strategies until a sufficient number of vaccines are available for more widespread, population-level coverage. Important work has been done to develop statistical methods to study these types of questions in studying, for example, the impact of cholera vaccinations within small family units[7], and extensions and refinements to a variety of complex settings[8] continue to be made.

---

[7] Perez-Heydrich et al. (2014) Assessing effects of cholera vaccination in the presence of interference. Biometrics. 70:731-744

[8] Benjamin-Chung et al. (2018) Spillover effects in epidemiology: parameters, study designs and methodological considerations. *International Journal of Epidemiology*, **47**:332-347