**The Data Science of Collective Action: Online Virality versus Real-World Impact**

[SLIDE Intro] Hi, my name is Sara Vera and I am a Senior Data Analyst at Causes.com. Causes is the world's largest campaigning platform. We connect people who support a common cause and empower them to take action together. In September, Causes launched the Supporter Network, an independent network that connects ideologically aligned individuals, celebrities, nonprofit organizations and socially responsible brands around the world to inspire collective action.

Campaigning is dependent on networking and peer-to-peer influence. The Internet could be the greatest organizing tool humanity has devised, connecting people from all parts of the globe and increasing the feeling that an individual is able to mobilize a group to accomplish group-based goals, which, in collective action literature, is called *perceived efficacy*. We've seen glimmers [Arab Spring SLIDE] of the Internet's potential to facilitate collective action in movements like the Arab Spring.

At Causes, the data we collect makes it possible for us to deliver the best campaigning tools suited to organizers' needs.

Today, I will discuss the data science of collective action and how we use data to inform a product that transforms [SLIDE about slacktivisim] "Slacktivism" to Activism. Slacktivism -- for those of you who don't know -- is a term to describe clicking buttons to feel like you're making a difference in the world. While the term "slacktivist" makes it sound like online activism is a feeble pretense for making a difference, in reality, leveraging online tools to maximize virality and get a message beyond "the choir" is quite powerful. Further, according to Sortable, "slacktivists" are twice as likely as the general population to volunteer, take part in an event, and ask for donations. What if organizers could better mobilize these online activists? Online networks help us leverage peer-to-peer influence and give power to the collective voice. [SLIDE Obama campaign] Even if you don't love Barack Obama, you can't deny that he squeezed every insight and action out of technology and data to make a successful bid for the presidency.

[CAUSES SLIDE] How can people optimize their campaigns and grow a campaign to scale without the tools and knowledge of content marketers? The answer is - our tools help people achieve this because they are built to leverage back-end data that helps ensure the right campaign is put in front of the right person, at the right time.

[SLIDE In this presentation] In this presentation, I will answer:

1.  Who are we building this for?
2.  Where does Causes data come from?
3.  How do we use data to inspire collective action?
4.  And how do we measure impact?

1

In answering these questions, I will help define how today's online influencers are using today's technology to better organize around common causes and campaigns for real-world change.

[SLIDE Who are our users] At Causes, our first mission is to get people civically engaged. We understand that to make a real-world impact, inspiring action is key. These actions are everything from reading and sharing content to starting a grassroots campaign.

[Good Ol' Clustering SLIDE] But first, to get people involved, we need to understand who they are. When we set out to build the Supporter Network, we started by performing a cluster analysis of our existing users to inform the direction that our product would take.

For this research, we subset our data to 200,000 users who used Causes between June and December 2012. From 230 variables, we derived 30 predictive variables by which to start categorizing our users, such as age, income, education, their activity level and topical campaign interests.

We ran k-means clustering algorithm on our entire data set, which resulted in 6 distinct personality types.

[K-means SLIDE] This is a visualization of our clustering work. One of our groups is excluded because it is essentially a small "other" category of people who don't really fit into any one cluster. You can see that the bottom two clusters on the slide exhibit drastically different behavior than the top three clusters--that's why they're "farther" away. While it's hard to gauge what distances like this really mean, but it's an informative visualization of what our user base looks like.

After performing the k-means cluster analysis to figure out who belongs in which cluster, we used these 6 clusters to train the [Random Forest SLIDE] random forest model to find out how predictive each variable is in identifying which cluster someone belongs to. Using the random forest model, we found 30 variables that best predicted user classification.

[Survey and Interview SLIDE] Using the behavioral and demographic trends we saw through our machine learning classification, we dug deeper into the motivations of our users through surveys and interviews. We received almost 1,500 responses to our 20 question survey.

[Results SLIDE] 98% of our users fit into one of these online activist personas.

*"The ambitious activist"* is in his mid-forties, and he's passionate, talkative, and enthusiastic about sharing his newfound passion for creating impact in the world with as many people as possible.

*"The practical activist"* is in his late thirties, he has focused ideas about how he can best affect change in the world and he seeks to find a soapbox where he can share his ideas with an audience. Sites like Causes are supplements to his offline activism, not a substitute.

*"The self-assured millennial"* is in his mid-twenties. He is self-confident and believes that he can play an integral part in changing the world around him. But it can be difficult for him to pledge allegiance to just one campaign or organization because if he sees injustice, he wants to get involved, no matter the context.

*"The organized retiree"* has recently retired after her a successful career and is not ready to slow down. As a practiced organizer, she approaches her work methodically: she researches legislation, educates herself and stays active through leadership roles in local organizations.

Although *"The tenacious veteran activist"* is retired and has a couple health problems that keep her mostly homebound, her enthusiasm for change is as prevalent as ever. She's been involved in campaigns for social change throughout her life; a lack of mobility has driven her to participate now online.

And while *"The casual participant"* would by no means identify herself as an activist, she does recognize the flaws and frustrations in the world around her. She visits sites like Causes through invitations from her friends, but does not feel much loyalty to the campaigns in which she participates.

[SLIDE Next iteration of data collection]

Once we understood our users, we used the data to inform tools that would help connect users and inspire peer-to-peer sharing. Just like a field office asks people to knock on 100 doors, or make 100 calls, we used our data to build a Supporter Network that would distribute the campaign responsibilities by connecting "*the casual participant"* with "*the tenacious veteran activist"* who share a passion for health education or the *self assured millennial* with the *organized retiree* who are concerned with environmental conservation.

So how do we know what these people care about?

[SLIDE] **Where does our data come from?**
There are several components to the Causes website where we gather most of our data. We collect demographic data from profile pages and campaign pages, and we collect a lot of

3

behavioral data when a user clicks or performs an action on our site. We also have a lot of offline data that we get from interviewing and surveying our users.

Here are some examples of Profile and Campaign pages to give you sense of the information that from them that we can use.

**Profiles pages and Campaigns:**
[SLIDE of my personal profile] Here is my Personal Profile page. You can see that I have displayed my civic identity by choosing virtual bumper stickers. Each sticker represents an issue that I care about, but is also a piece of data that we use at Causes to make decisions about our product and the best way we can facilitate supporter connections and campaign success.

Over here, you can see whom I support and who supports me. Supporting basically means, "Hey, I like what you stand for, let me know how I can help." Who you support is meant to be very intentional. Using the supporter network -- individuals, nonprofits and brands can reach an interested audience and draw attention to specific issues and campaigns. We also use the supporter network to analyze demographics related to campaigning, fundraising, and political affiliations. Don't worry, only the data scientists know how to get this information. You can display it, but no one has access to this information besides me, basically.

[Brand SLIDE] Brand and [Org SLIDE] Organization profile pages are a similar setup. When an organizer creates a campaign, they have the option of "tagging" the campaign with relevant issue categories so people can easily search for it. [Sierra Club Campaign Page] For example, Sierra Club campaigns might be tagged with 'conservation,' 'environment,' and more descriptive, specific tags depending on the campaign, such as 'oil and gas conservation.' This tagging system also allows us at Causes to categorize the campaigns on our website into "issue" pages in order to surface relevant content to our users. It is also possible to post comments, stories and photos, giving people a way to become more engaged and personally involved.

[Tony SLIDE] And finally, we have an example of a personal campaign page, which is a way for a supporter to bite off a smaller piece of a larger campaign to make the overall goal more tangible. Tony is collecting 10 signatures for a larger campaign that is collecting a total of 100 signatures for a petition to make North Beach sidewalks more pedestrian friendly. If Tony invites you to sign this petition, you can click back to the `overall campaign to read more about the San Francisco County Transportation Authority, the neighborhood studies being conducted and other specific background information regarding this petition.

So I just walked you through how people set up their civic identity, create a supporter network, and take action on campaigns they care about. From here, we use this information to make this process more efficient and engaging.

[SLIDE Recommendation Engines] Some would say you need heart and passion but here, we need Discovery!

4

Discovery is a huge first step to engaging our users. And with all of the information we see on the profile pages, the virtual bumper stickers, campaign issue tags, and behavioral data, we try to surface relevant content and help users determine who in their existing online networks on Facebook, LinkedIn and Twitter are likely to share their campaign interests.

While users are building their supporter networks and taking action, we continue to refine and suggest supporter-connections, issues you might find interesting, and inspire you to take action.

This leads to the obvious question of assessing whether your actions have an impact in the real world.

[SLIDE] **Measuring Impact**

Measuring impact is a perennial problem for organizations invested in social change. It's hard enough to define impact, let alone figure out metrics to keep track of. People need to feel that taking action online is a more rewarding experience than sending a tweet or posting on Facebook. If someone starts or participates in a campaign, is there a return on their investment? Can we use our data to show the impact of these campaigns?

[SLIDE Theory of Change] There has been some research from the Harvard Business School about how to measure impact. In this line of research, the authors outline a contingency framework for measuring results based on the causal logic of change that underlies any intervention program.

The logic chain includes an organization's Inputs and Activities that lead to outputs, outcomes, and ultimately, impact.

Inputs include funds, equipment and supplies, knowledge and technical expertise

Activities include basic needs delivered, such as food and shelter; or services delivered, such as training programs

Outputs, are the immediate results

Outcomes are the medium-term results such as improved living conditions and health, increased incomes, and enhanced political voice

5

And Impacts are long-term results that indicate fundamental changes in social norms

Deciding what to measure depends on where in the logic chain your company or group lands. Given the diverse mission and capacities of companies and organizations, some should be monitoring long-term impacts, while others should stick to reporting immediate results.

At Causes today, [Causal Logic SLIDE] we have data to measure short-term outcomes in the first half of this logic chain. We keep track of money raised, petitions signed and other shorter-term outcomes of campaigns that our site facilitates.

Here's an example.

[Elephant SLIDE] You may have heard the controversy over an NRA-sponsored hunting show on the NBC Sports Network last week. The leader of this campaign was outraged when he saw an NRA lobbyist shoot a threatened bull elephant in the face on "Under Wild Skies". The leader of this campaign originally created a petition to NBC to stop airing this episode to a national audience. He asked, "Do the values of NBC really fall in line with this type of programming where endangered animals are hunted down by the world's leading pro-gun lobby?"

This leader was able to use our supporter network to identify a large of group of people concerned with animal rights, making participation and sharing rates for the petition very high. This is part of getting the right content in front of the right people, at the right time.

[Personal Campaigns SLIDE] The momentum really got going when 639 Personal campaigns were started to help the leader reach his overall goal of 100,000 signatures.

Here, the transition from clicking to creating is the difference between slacktivism and online activism.

Then, [Kirstin Davis SLIDE] glamorous celebrities like Kirstin Davis from Sex and the City started re-tweeting the petition, then [Media SLIDES] the LA Times, Huffington Post and other media outlets picked up the story, increasing public pressure.

[Elephant SLIDE] In the end, the collective influence of over 115,000 people helped to pressure NBC to cancel the program.

This is how online action can be more effective than actual feet hitting the pavement. Participation is democratized in a way that inspires over 100,000 people to sign a petition within 4 days. As users see the number of people involved increasing, they are even more inspired to act. Digitizing our civic engagement allows everyone to get involved and say, "Hey, I stand behind this cause."

[Logic Chain SLIDE] So, again, at Causes, here we are in the first part of the logic chain where we see the immediate result of this petition -- cancelling this NRA-sponsored television show.

We can leverage and display success by showing how many people have signed the petition and created their own personal campaigns to bite off a piece of the larger goal.
But this doesn't measure the long-term impact.

[Outcomes SLIDE] There are several outcomes of this campaign. Causes was able to connect animal rights advocates concerned with the brutality of the hunt in which the elephant is not immediately killed, as well as gun control advocates who questioned the promoting of an NRA-sponsored program to a national audience. In addition to getting signatures, the campaign pages allowed supporters to collaborate and helped to raise awareness and recruit additional support.

But in the future, we'd like to say that this petition contributed to an overall societal change in values.

[Data and Impact SLIDE] These longer-term outcomes are both challenging and exciting for data scientists. The ability to use data about who we are building for to create tools for short-term results is a promising start to measuring long-term impact.

We need to figure out indicators of societal change that we can measure, and take note of how many campaigns of what types and volume and frequency it takes before we reach some threshold of "change." It about makes my head explode just thinking about it!

But...

The more we can shape and refine strategy based on people's behavior and practices, the more likely we will have lasting impact.

[THANK YOU SLIDE]