

“Previsão imediata” de danos por desastres



Do “paradoxo da amizade” à avaliação do impacto de um furacão,
usando dados do Twitter

MANUEL CEBRIÁN
Sociólogo da informática

A história em que este projeto se baseia começou em 2010, quando eu estava na UC San Diego. Era o início da pesquisa usando conjuntos de dados de redes sociais. Para um jovem bolsista como eu, a UCSD era um lugar fascinante para começar a fazer esse tipo de trabalho. James Fowler, uma das mentes mais brilhantes na minha área, a sociologia da informação, trabalhava no campus. James e Nicholas Christakis – conhecidos como a “dupla dinâmica” que descobriu a

ligação entre obesidade e relações sociais – publicaram um dos meus documentos favoritos neste campo. Nele eles demonstraram que o paradoxo da amizade – o fenômeno em razão do qual seus amigos têm mais amigos que você, descoberto por Scott Feld em 1991 – poderia ser usado para detectar surtos de sarampo antes do que se conseguiria apenas observando a população em geral.

No melhor dos casos, os métodos atuais para detecção de focos contagiosos fornecem informações contemporâneas sobre o desenvolvimento de uma epidemia. Sabe-se que, durante um surto, as pessoas situadas perto do centro de uma rede social têm em média maior probabilidade de se infectar antes que aquelas que estão na periferia. Infelizmente, mapear indivíduos localizados em uma posição central que poderiam ser monitorados para se observar a evolução da infecção costuma ser muito difícil. O que propomos é uma estratégia alternativa que não exige a confirmação de uma estrutura de rede global, ou seja, basta apenas monitorar os amigos dos indivíduos selecionados aleatoriamente. Esses indivíduos são conhecidos por ocupar uma posição mais central. Para avaliar se um grupo de amigos dessas características permitiria a detecção precoce, estudamos um surto de gripe que ocorreu na Faculdade de Harvard no final de 2009 e acompanhamos 744 alunos que eram membros de um grupo de indivíduos escolhidos ao acaso ou de um grupo de seus respectivos amigos. A partir de diagnósticos clínicos, a evolução da epidemia no grupo de amigos se deu 13,9 dias antes (95% C.I. 9,9-16,6) da evolução no grupo escolhido ao acaso (ou seja, a população considerada como um todo). Além disso, o grupo de amigos mostrou um tempo de diagnóstico significativo ($p < 0,05$) no dia 16 da epidemia, 46 dias antes do pico de incidência diária na população contemplada como um todo. Este método usado como sensor poderia fornecer um tempo adicional significativo para se reagir a epidemias em populações pequenas ou grandes submetidas a vigilância.

Christakis NA, Fowler JH (2010)

Sensores em redes sociais para a detecção precoce de surtos contagiosos.

PLoS ONE 5(9): e12948. doi:10.1371/journal.pone.0012948

Durante um tempo eu fiquei bastante interessado em testar o comportamento deste “paradoxo da amizade” nas novas estruturas sociais que as redes sociais digitais ofereciam. Esteban Moro (da UC3M, Espanha) estava visitando a UC de San Diego no verão e, junto com outro colega espanhol que estava no campus, Manuel García-Herranz (da UAM, Espanha), começamos a “dissecar” alguns dados do Twitter para examiná-los. Os resultados preliminares eram promissores: o paradoxo da amizade também se refletia no Twitter. Mas nós nos perguntamos o que aquilo significava... Será que era outra versão do fenômeno do paradoxo da amizade na

Internet? Decidimos falar com James e Nicholas. Eles sugeriram que, da mesma forma que a gripe se espalhou de uma pessoa para outra no campus da Harvard, a informação também se espalhava online, e talvez o paradoxo da amizade também poderia ser usado para detectar um contágio global na Internet. Arregaçamos as mangas para investigar esta questão e, após muitas horas de trabalho, conseguimos demonstrar que era isso mesmo, que também na Internet havia a possibilidade de usar os amigos como sensores para detectar surtos contagiosos globais/em escala.

Recentemente, foram conduzidas pesquisas focadas no monitoramento de dados online em escala global para melhorar a detecção de epidemias, padrões de humor, movimentos no mercado acionário, revoluções políticas, faturamento com bilheteria e comportamentos dos consumidores e muitos outros fenômenos importantes. No entanto, as considerações sobre privacidade e a grande amplitude dos dados disponíveis online fazem com que o monitoramento global rapidamente se torne inviável e os métodos existentes não aproveitam ao máximo a estrutura de rede local para identificar os nós críticos na realização do monitoramento. Nós, então, desenvolvemos um modelo da disseminação epidêmica da informação em escala global sobre uma rede social articulada publicamente e demonstramos que um método simples não só pode produzir uma detecção precoce como também pode alertar sobre a existência de focos infecciosos. Nesta abordagem, escolhemos aleatoriamente uma pequena fração de nós na rede e, em seguida, escolhemos aleatoriamente um amigo de cada nó para incluí-lo em um grupo para o monitoramento local. Utilizando dados de 6 meses procedentes principalmente da esfera do Twitter, demonstramos que este grupo de amigos tem uma posição mais central na rede e ele nos ajuda a detectar surtos virais relativos ao uso de hashtags novos aproximadamente 7 dias antes do que conseguiríamos fazer usando um grupo de mesmo tamanho escolhido ao acaso. Além disso, o método realmente funciona melhor que o esperado devido à estrutura da rede, já que os atores que ocupam uma posição marcadamente central são mais ativos e mostram maior diversidade nas informações que transmitem aos outros. Estes resultados sugerem que o monitoramento local não é apenas mais eficiente, mas também mais eficaz e mostram que pode ser aplicado no controle de processos contagiosos em redes em escala global.

Garcia-Herranz M, Moro E, Cebrian M, Christakis NA, Fowler JH (2014)

Usar os amigos como sensores para detectar surtos contagiosos em escala global.

PLoS ONE 9(4): e92413. doi:10.1371/journal.pone.0092413

Esta descoberta produziu um número significativo de pesquisas acadêmicas sobre as aplicações online do paradoxo da amizade (como no caso de Kristina Lerman et al., da Universidade do Sul da Califórnia, Kay Axhausen et al., na ETH Zúrique e Y.H Eom et al., na Universidade de Toulouse). Seus estudos, como se deu com o nosso, mostraram que eles conseguiam detectar a disseminação viral na Internet, a “disseminação da informação”, mas o que acontece no mundo real? Podemos usar o paradoxo da amizade para detectar algo que está acontecendo no mundo físico?

Esta pesquisa foi retomada em 2012, quando eu estava no meu novo cargo no NICTA (hoje Data61, após a fusão do NICTA com o CSIRO), o centro mais importante da Austrália em pesquisa sobre ciência dos dados. Trabalhei lá com Pascal Van Hentenryck (que hoje está na Universidade de Michigan), especialista em gestão e mitigação de desastres em tempo real. Pascal me recrutou para que trabalharmos juntos e explorar se as

informações em tempo real produzidas pelas redes sociais eram informativas o suficiente a ponto de melhorar os planos de computação de evacuação em que vínhamos trabalhando há anos. Enquanto analisávamos isso, o furacão Sandy atingiu os Estados Unidos e, vendo a tremenda devastação que ele causou, nós nos perguntamos: Será que podemos usar o paradoxo da amizade para detectar precocemente a destruição que está ocorrendo na costa leste dos Estados Unidos? Para esta missão, contamos com dois associados com pós-doutorado e muito talento: Caron Chen, recentemente doutorada em GIS (geoprocessamento), que se dedicou a obter o máximo de dados possível sobre o Furacão Sandy no Twitter, e Yury Kryvasheyev, recentemente doutorado em Física (modelos computacionais), que se concentrou em verificar se as pessoas mais conectadas enviariam um Tuíte antes que os outros quando o furacão os atingisse. E, mais uma vez, funcionou: os sensores de redes sociais conseguiram enviar alertas precoces durante o impacto do furacão Sandy:

O fluxo de informações durante eventos catastróficos é um aspecto crucial da gestão de desastres. As plataformas modernas de comunicação, particularmente as redes sociais online, proporcionam uma oportunidade de estudar esse fluxo e extrair sensores de alerta precoce, melhorando, assim, o grau de preparação e de resposta a emergências. Neste caso, estuda-se o desempenho do método de sensores das redes sociais, que se baseia em propriedades topológicas e propriedades de comportamento decorrentes do “paradoxo da amizade”, observando-se mais de 50 milhões de mensagens no Twitter publicadas antes, durante e depois do furacão Sandy. Verificamos que as diferenças na centralidade da rede dos usuários realmente significa uma vantagem de ciência moderada (até 26 horas) e que a localização geográfica dos usuários dentro ou fora da área afetada pelo furacão desempenha um papel significativo na determinação da escala dessa vantagem. A resposta emocional parece ser universal, independente da posição que ocupe na topologia da rede e mostra padrões característicos e facilmente detectáveis, o que abre a possibilidade de se implementar uma técnica simples de “percepção do sentimento”, que pode detectar e localizar desastres.

Kryvasheyeu Y, Chen H, Moro E, Van Hentenryck P, Cebrian M (2015)

Desempenho dos sensores de rede social durante o furacão Sandy

PLoS ONE 10(2): e0117288. doi:10.1371/journal.pone.0117288

Naquela época (2014–2015) o Twitter já era identificado como uma plataforma que vem a ser muito útil quando ocorrem desastres naturais ou tecnológicos. Os pesquisadores estudaram a fundo o papel dos serviços online por facilitarem o fluxo de informações relacionadas com as emergências, detectando eventos online e arrecadando fundos de forma colaborativa para diminuir os danos. Mais uma vez, a maioria dos estudos até o momento (Guan e Chen, 2014, Uso de dados de redes sociais para compreender e avaliar os desastres), com raras exceções, se relacionava à informação e juntos deram origem à área que hoje se conhece como informática para crises. Agora, o próximo passo consistia não só em detectar informações sobre o desastre como também em determinar se poderíamos inferir a magnitude dos danos que estavam ocorrendo nas diversas comunidades usando exclusivamente os dados do Twitter. Nosso passo seguinte foi uma aposta arriscada, já que achávamos que iria muito além do que podíamos tirar do Twitter. Uma coisa é a informação, em relação à qual sabemos que as redes sociais se comportam bem, mas e os danos no mundo real? É um nível de impacto muito maior e não tínhamos certeza de que todos os vínculos estatísticos poderiam sobreviver ao ruído presente no Twitter.

No decorrer da análise, portanto, estabelecemos a distribuição espacial dos tuítes durante a supertempestade Sandy. Os resultados preliminares de Yuri indicaram que a atividade online normalizada (mensagens políticas sobre o tema per capita) tem uma relação direta com a proximidade da zona do desastre, com um aumento acentuado nas proximidades do ponto de impacto. Isto nos deu esperança de que a força motriz subjacente ao tuíte fosse uma combinação de fatores que fazem com que o fato de tuitar seja relevante: até que ponto o evento é perigoso, qual seu nível de proximidade, até que ponto suas manifestações são óbvias (ventos, ondas de maré ciclônicas, etc.) e qual é o nível de severidade das consequências? Acreditamos que se poderia detectar a relação entre a atividade virtual e os danos causados no mundo real.



As medições dos danos no mundo real são difíceis de obter. Felizmente para nós, Yury Kryvasheyev entrou em contato com um especialista em danos causados por desastres. Era Colleen O’Dea (do [NJ Spotlight](#)), que nos ajudou muito com orientações sobre como obter dados relacionados aos danos causados pelo furacão Sandy de diferentes agências (FEMA, o Departamento de Banking e Seguros do Estado de Nova Jersey e o Departamento de Serviços Financeiros do Estado de Nova York).



Posteriormente, decidimos incluir antigos membros da equipe (Esteban Moro, James Fowler) para que nos ajudassem a avaliar as análises estatísticas que consideramos mais graves do que as que apareciam em documentos anteriores. Para processar os dados, James também conseguiu a colaboração de um jovem cientista climático, Nick Obradovich, que estava fazendo doutorado na Instituição Oceanográfica Scripps.

Com todos os ingredientes à mão, começamos a comparar duas distribuições espaciais: a atividade online normalizada e os danos causados pelo furacão per capita. E, comprovando nossa suposição intuitiva, a correlação foi bastante significativa (alcançou um valor de 0,6 na correlação de Pearson). Os resultados foram graduados: por exemplo, em relação aos danos, a atividade posterior ao desastre foi mais preditiva que a enxurrada de tuítes durante o pico do desastre. Da mesma forma, uma resolução espacial da análise afeta a força das correlações observadas. Por fim, a metodologia da avaliação de danos (que se considerou) e as subvenções individualizadas da FEMA estão moderadamente correlacionadas com a atividade online, mas em conjunto com os dados sobre o seguro, o relacionamento é muito mais forte (sugerindo que, quanto mais completa a imagem do dano monetário, mais estável é a relação). Fundamentalmente, verificamos nossas conclusões em relação a outros eventos ocorridos nos Estados Unidos e encontramos resultados semelhantes em relação às escalas espaciais e tipos de ameaças: grandes tempestades e tornados, inundações, deslizamentos de terra e terremotos.

Em suma, deixando de lado os detalhes mais elaborados, o resultado mais importante é uma solução muito animadora de custo praticamente zero que permite fazer a “previsão imediata” dos danos causados por desastres. E não é só isso: nosso método também poderia ajudar na avaliação posterior da catástrofe, ajustando os relatórios oficiais e as ferramentas preditivas (com base, por exemplo, na trajetória do furacão e nas simulações de fragilidade) aos dados das redes sociais e avaliando o impacto, a gestão e a eficácia das medidas adotadas no decorrer do desastre. Com o monitoramento do furacão usando as redes sociais, os gestores de emergência podem adotar ações nas áreas em que as redes sociais mostraram uma atividade anormalmente alta. Em alguns países menos equipados para lidar com desastres, há um interesse considerável em explorar as redes sociais no que diz respeito à resposta a emergências. Nossos estudos representam passos positivos nesta direção, pois indicam que as redes sociais captam alguns elementos que são difíceis de obter com os meios tradicionais. ■