



Methods for quantifying effects of social unrest using credit card transaction data

Xiaowen Dong^{1*} , Joachim Meyer², Erez Shmueli², Burçin Bozkaya³ and Alex Pentland¹

*Correspondence: xdong@mit.edu

¹ MIT Media Lab, Cambridge, USA

Full list of author information is available at the end of the article

Abstract

Societal unrest and similar events are important for societies, but it is often difficult to quantify their effects on individuals, hindering a timely and effective policy-making in emergencies and in particular localized social shocks such as protests. Traditionally, effects are assessed through economic indicators or surveys with relatively low temporal and spatial resolutions. In this work, we compute two behavioral indexes, based on the use of credit card transaction data, for measuring the economic effects of a series of protests on consumer actions and personal consumption. Using data from a metropolitan area in an OECD country, we show that protests affect consumers' shopping frequency and spending, but in noticeably different ways. The effects show strong temporal and spatial patterns, vary between neighborhoods and customers of different socio-demographical characteristics as well as between merchants of different categories, and suggest interesting subtleties in purchase behavior such as displaced or delayed shopping activities. Our method can generally serve for the real-time monitoring of the effects of major social shocks or events on urban economy and consumer sentiment, providing high-resolution and cost-effective measurement tools to complement traditional economic indicators.

Keywords: Social shocks; Economic effect; Consumer behavior; Spatiotemporal pattern; Credit card transaction

1 Introduction

The routine of daily life is punctuated by extraordinary public events, ranging from major sports and cultural events, to terror attacks or disasters. Some events are local, and some are national or even wider; some affect only part of the population, while others affect the entire population. Although unlike natural events and disasters, certain social events, such as a protest, a riot, a large police action, or a bomb explosion can still have characteristics of natural emergencies, with fatalities and damage to properties. However, when such localized social shock occurs, we have mostly qualitative or at best retrospective survey data about its effect on the surrounding community. Consequently it is impossible to know how well authorities managed the side-effects of the event, and thus difficult to develop effective event policies for use by police and municipal authorities.

In particular, understanding the effects of major social shocks or events on the economy and consumer behavior can have important implications [1–7]. Existing measures of economic behavior and attitudes towards economic situations use surveys to assess consumer confidence or consumer sentiment, asking about the intentions to purchase goods or to

make large investments. These include objective measures published by national agencies, such as the measure of personal consumption expenditures in the monthly personal income and outlays report, provided by The Bureau of Economic Analysis (BEA) [8]. Other indicators are based on consumers' subjective reports, such as the monthly consumer confidence index (CCI) issued by The Conference Board [9], and the monthly index of consumer sentiment (ICS) published by the University of Michigan [10]. These data and the actual economic behavior, as expressed in consumption, can be strongly correlated [11], and they can therefore be used to evaluate the effects of events. However, their temporal resolution is low (consumer confidence and sentiment surveys are usually conducted on a monthly base). Their spatial resolution is low, too, since the measures aggregate information over large areas and usually entire economies. Finally, the information usually cannot be used to look at the differential effect of events on different segments of a society.

The limitations of traditional approaches may be overcome by using digital tools to study human and social behavior; analyzing large-scale quantitative data in emerging fields such as computational social science [12, 13]. The analysis of large amount of data accumulated in social networks, cellular phone records, or credit card transactions may allow us to gain a new understanding of social processes. Such data have been widely utilized to study situation awareness and response to emergencies [14–18], mostly due to natural events and disasters, as well as the dynamics of communication and information propagation immediately followed [19–22]. However, relatively little work, so far, aimed to understand the impact of social events, and especially localized social shocks, using quantitative behavioral data.

Einav et al. recently highlighted the new possibilities of conducting economic research in the age of “Big Data” [23]. Within this scope, in this work, we show how a particular type of data, namely credit card transaction records, can be used to quantify and understand the repercussions localized social events have on individuals' economic behavior, quantifying when, where and how an event has an effect. In particular, credit card data help us overcome the limitations of traditional indicators by allowing us to measure purchase behavior directly, and, together with information about the types of merchants at which purchases are made, to infer what kinds of products are purchased. Especially, purchases at physical shops (rather than those made online) are clearly linked to specific spatial locations and time stamps. Furthermore, credit card issuers have information about the demographics and economic status of the credit card holder, which enables us to measure the effects of events on different demographic groups in the population. This would allow the authorities to develop targeted and timely policies in case of emergencies.

2 Data and methods

To demonstrate the use of this method, we analyzed more than ten million credit card transaction records provided by a major financial institution in an OECD country about more than 100 thousand individuals, at more than 100 thousand merchants during a period of three months. Each record consists of the date, time and amount (in local currency) for one credit card transaction, along with anonymized customer and store IDs. Additional information about the customers and stores is also available, such as customers' gender and the neighborhoods in which they live, as well as the category of the stores and the neighborhoods in which they are located.

We focus our analysis on credit card transactions, made at stores located in the largest metropolitan area of the country. We omitted foreign and online transactions as well as

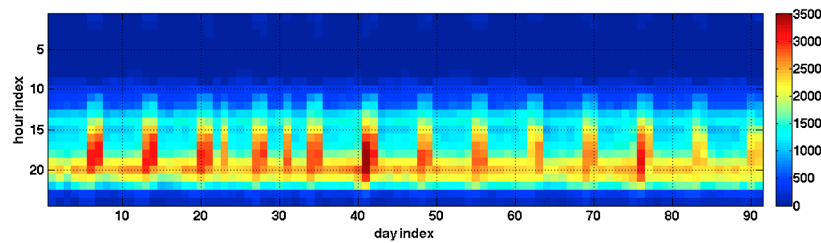


Figure 1 Visualization of the hourly number of transactions in the metropolitan area during the period the data set covers. The horizontal axis corresponds to the index of the days in the data set, and the vertical axis corresponds to the index of the specific hours of each day

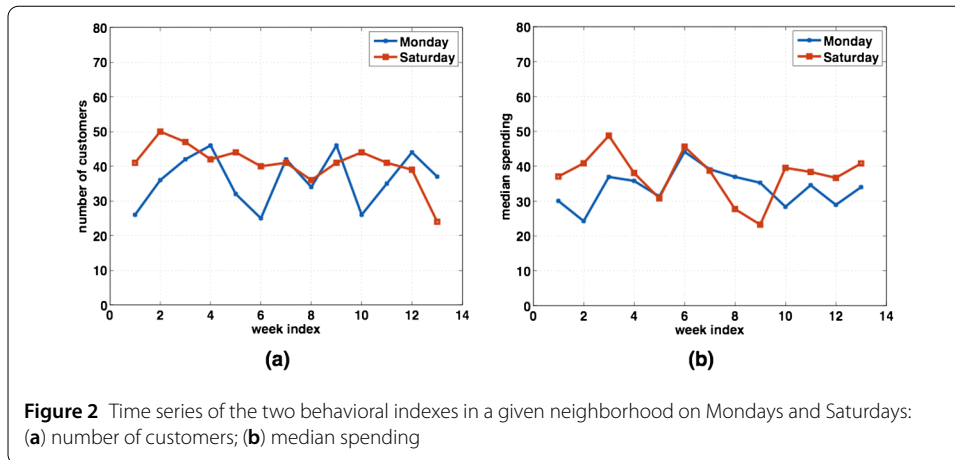
transactions without timestamp, stores with fewer than ten distinct customers over the three-month period, and customers who visited fewer than ten distinct stores.^a As we are mainly interested in people selectively changing their purchase behavior in physical locations due to social events, we further selected five broad merchant categories that mostly correspond to on-site and discretionary purchases, i.e., “amusement and entertainment”, “clothing stores”, “retail stores” (including subcategory “grocery stores, supermarkets”), “personal service providers”, and “miscellaneous stores” (including subcategory “eating places, restaurants”). The resulting data set covers 1.8 million records from 100 thousand customers at 18 thousand geo-localized stores in the metropolitan area. Figure 1 visualizes the hourly evolution of the number of transactions made by the customers, from which we can already observe strong weekly patterns that separate the purchasing behaviors on weekdays and weekends, and daily patterns that distinguish different periods of a day such as the morning, afternoon and evening.

We quantify the impact of localized social events on purchase behavior at the level of administrative neighborhoods in the metropolitan area. For each neighborhood, we look at the transactions at the stores inside the neighborhood. We then compute, on a daily basis, two behavioral indexes, using the credit card transaction data:

1. *consumer action*: the number of unique customers who made transactions at stores inside the neighborhood;
2. *personal consumption*: the median spending amount in local currency for all transactions inside the neighborhood.

The first index is a proxy of the number of customers shopping in the area. There often exist many alternatives for places to shop, so the number of people choosing a particular area is a proxy for their preference toward this area versus other areas. This therefore serves as a proxy for how comfortable people feel when visiting the area. The second index is a proxy for the individual’s consumption expenditure, which measures the consumers’ mood and the level of economic security they feel, e.g., whether they feel that they should delay certain purchases or save money due to social conditions.

For each day in the week and for each neighborhood in the metropolitan area, we construct a time series with values for the two indexes. The data cover a period of 13 weeks. We therefore obtain two time series, each with 13 data points for each day in the week, for the number of customers and the median spending amount in a neighborhood on a particular weekday or weekend. For example, Fig. 2(a) and Fig. 2(b) show the time series of the number of customers shopping, and the median spending amount, respectively, in a given neighborhood on Mondays and Saturdays. To normalize differences in shopping



behaviors between different days of the week (a difference that is particularly noticeable between weekdays and weekends), we compute the relative deviation from the mean,^b which is defined as:

$$c = (x - \bar{X})/\bar{X}, \quad (1)$$

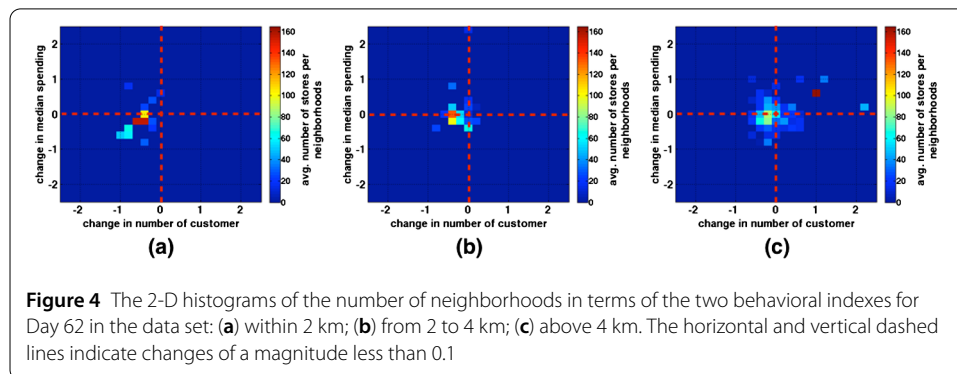
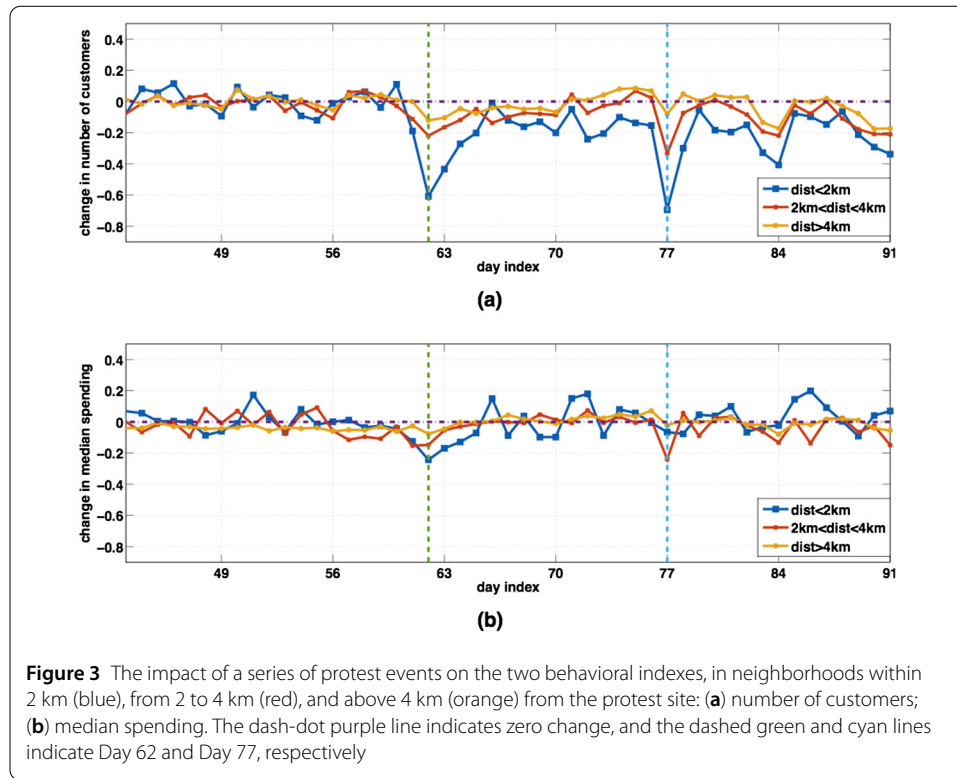
where x is the value of an index on a given day, and \bar{X} is the mean value of the index for the given day during the first six weeks of the data. We take the first six weeks as the reference period, since there were no particularly noteworthy events during this period.^c We obtained two scalar values for each neighborhood for each day, which indicate the temporal evolution of the two behavioral indexes for each weekday or weekend.

3 Results

We use the score of relative deviation from the mean for the two behavioral indexes, computed in each neighborhood for each day in the data set, to measure the impact of a series of protests that took place in a central site of the metropolitan area during a period of one month. In particular, we focus on the temporal, spatial, heterogeneous, and integral economic effect of these protest events.

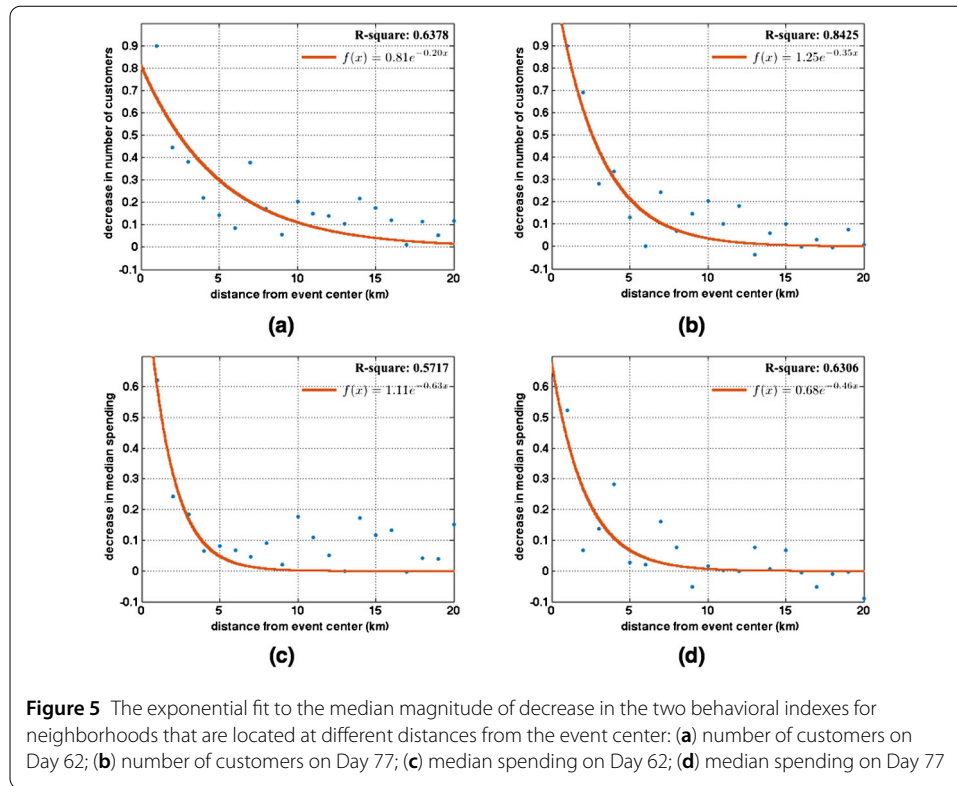
3.1 Temporal variation of effects

In Fig. 3(a), (b), for more than 300 neighborhoods that have an average number of customers above ten in our data set, we show the median (weighted by number of stores in the neighborhoods) change in the two indexes, both temporally and spatially. More specifically, daily fluctuations of the changes in behavioral indexes are illustrated with respect to a reference period of six weeks, namely, from Day 1 to Day 42 in our data set, prior to the beginning of the societal unrest. Two major events of protests occurred on Day 62 (highlighted in green) and Day 77 (highlighted in cyan). The number of customers on these days decreased sharply in the vicinity (within 2 km) of the protest site, but less drastically in the neighborhoods within 2 to 4 km distance from the protest site and even less at larger distances. After the initial decrease, median spending returned more quickly to normal, and it was not as strongly affected as the number of customers. Here, too, the effect was generally stronger near the protest site.



To confirm our findings, we test the statistical significance of the results shown in Fig. 3(a), (b) in the Appendix (see Table A3(a)). On the days of two major events, we see a significant drop in both behavioral indexes for all three distances in Fig. 3(a), (b). In addition, there also exist statistically significant differences between distances. On days following the major protests (those after Day 62 and Day 77), the number of customers decreased significantly in neighborhoods within 4 km from the protest site, but not beyond. On the other hand, there was no significant decrease in the median spending in any distance, suggesting that the people who went out shopping did not tend to spend less money than usual.

The combined effect of the events on the two independent behavioral indexes can be seen in 2-D histograms, such as the ones in Fig. 4(a), (b), (c). The horizontal and vertical axes represent the score of relative deviation from the mean for the two behavioral indexes, respectively, and the color code indicates the average number of stores per neighborhood



for neighborhoods that have a certain combination of the two scores. In such a histogram, positions towards the upper right represent increased numbers of customers and larger median spending, while positions towards the bottom left represent the opposite. If no change occurs, all neighborhoods stay at the center of the diagram. It is clear from Fig. 4(a) that on Day 62 stores in most neighborhoods were negatively affected by the protests, with fewer customers (mostly 30–90% less) and smaller spending amounts (mostly 10–70% less) within 2 km of the protest site (event center). People went out shopping less and were reluctant to spend as much as they usually do, possibly due to the unstable situation in the area. These trends were less pronounced when moving away from the event center, as shown in Fig. 4(b), (c), which is consistent with the results shown in Fig. 3(a), (b). The same 2-D histograms for Day 77 are shown in Fig. A1 in the Appendix.

3.2 Spatial decay of effects

To quantify the spatial propagation of the event effects, we study the effect as a function of the geographical distance from the event center. In Fig. 5, the blue dots in both figures indicate the median (weighted by number of stores in the neighborhoods) magnitude of decrease in (i) number of customers (Fig. 5(a), (b)) and (ii) median spending (Fig. 5(c), (d)) for neighborhoods that are located at different distances from the event center (i.e., the dot at 2 km represents all neighborhoods that are further than 1 km but less than 2 km away from the event center).^d Fitting the blue dots with exponential decay functions for both cases (shown as red curves), we see reasonably good fits, which indicates that the negative effect of events decayed approximately exponentially with respect to distance from the event center. We then compute the exponential decay constant, and its inverse, which is the so-called mean lifetime of the decay (mean distance in our context). The mean decay

distances for Day 62 and Day 77, based on Fig. 5(a), (b), were $1/0.20 = 5$ km and $1/0.35 = 2.86$ km, respectively, and, based on Fig. 5(c), (d), $1/0.63 = 1.59$ km and $1/0.46 = 2.17$ km, respectively.

These results indicate that, first, the amount of money people spent when making a purchase was less clearly affected by the protests, compared to the change in the number of customers, and the negative effect spread less further away from the event center as well. Second, it seems that the event on Day 77 had a stronger impact on consumer actions within the close vicinity of the event center, which is indicated by a higher initial quantity of 1.25 in the exponential fit and reflected by the magnitude of the first two points in Fig. 5(a), (b). Such effect, however, decayed spatially faster compared to Day 62, which is suggested by a smaller mean distance of 2.86 km and reflected by the generally smaller magnitude from the third point (3 km) onwards in Fig. 5(a), (b). On the other hand, on Day 62 personal consumption was more affected close to the event center, while the negative effect on this index spread further away on Day 77. Such differences might be due to people's responses to events changing over time. Finally, we can also study the event effect as a function of the estimated travel time (by car) from the event center, thus taking into account geographical constraints and transportation accessibility. The results are similar (in terms the exponentially decaying patterns) and presented in Fig. A2 in the Appendix. We further show in Fig. A3 and Fig. A4 in the Appendix the event effect on the individual neighborhoods. Overall, these results demonstrate the possibility of utilizing transaction data to quantify the spatial decay of event effects.

3.3 Heterogeneous effects

Social events may impact neighborhoods of different characteristics and the purchase behavior of people from different demographic groups in different ways. It may also have a differential effect on purchases from different types of merchants. We therefore study the heterogeneous effects of the social events in the following scenarios, by focusing on neighborhoods within 4 km from the event center.

3.3.1 Socio-economic status

We first show in Fig. 6 the time series of median (weighted by number of stores in the neighborhoods) change in the two behavioral indexes, computed over neighborhoods of higher socio-economic status, compared to that for those of lower socio-economic status. The socio-economic status is a composite measure between 0 and 100 that quantifies the relative prosperity of the neighborhood based on a number of indicators such as income and education level, which is obtained from the results of a recent census provided by the National Statistical Institute of the country. The higher the index, the more prosperous the neighborhood is. We see that the number of customers decreased more sharply in neighborhoods of lower socio-economic status for days following the protests. On the two days of major events, however, such a decrease was slightly higher in wealthier neighborhoods, possibly due to the fact that a larger portion of the demonstrators was from these neighborhoods. On the other hand, median spending dropped more significantly in less wealthy neighborhood on these two days, but not afterwards. Tests for statistical significance of the differences are presented in Table A3(b) in the Appendix. Overall, our results suggest that the events had longer lasting negative effects on less wealthy neighborhoods,

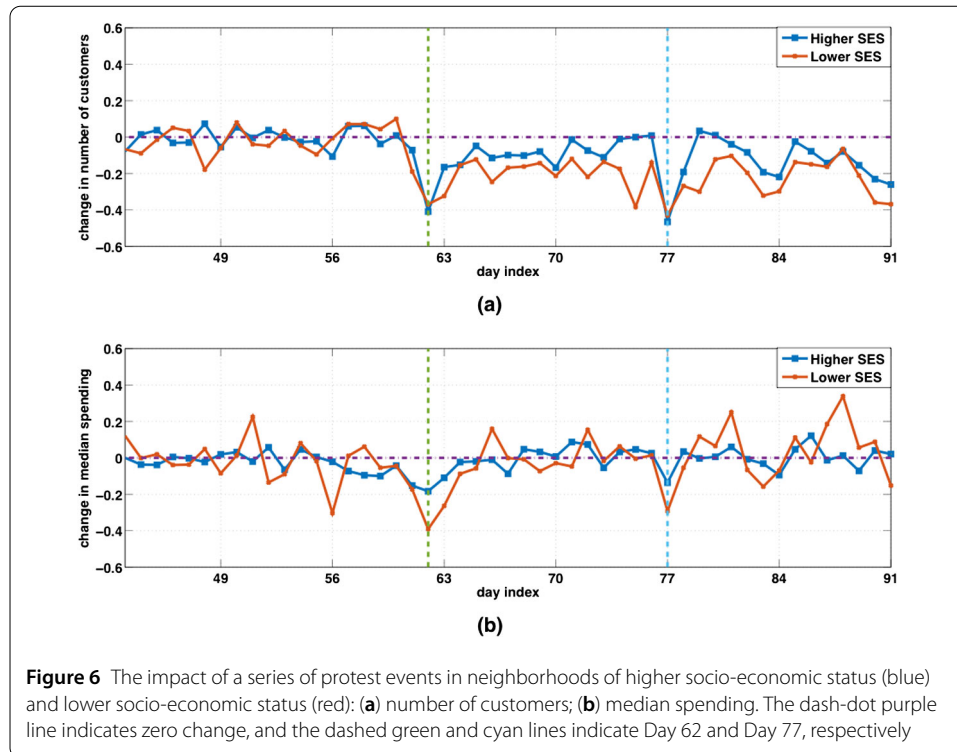


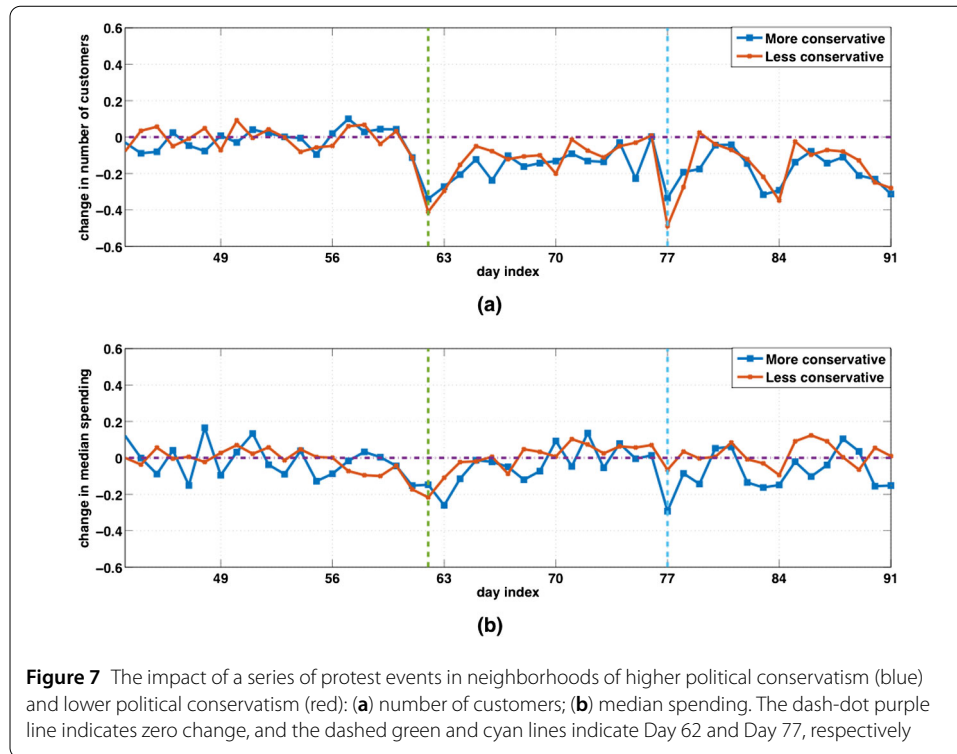
Figure 6 The impact of a series of protest events in neighborhoods of higher socio-economic status (blue) and lower socio-economic status (red): **(a)** number of customers; **(b)** median spending. The dash-dot purple line indicates zero change, and the dashed green and cyan lines indicate Day 62 and Day 77, respectively

but mainly in terms of consumer action rather than personal consumption. Because political preference is shown to be correlated with wealth in the same country [24], we next examine the effects on neighborhoods distinguished by political conservatism.

3.3.2 Political preference

Figure 7 shows the same time series as in Fig. 6, computed over neighborhoods of higher political conservatism, compared to that for those of lower political conservatism. An index for political conservatism of each neighborhood is a measure between 0 and 100 that is computed based on the percentages of votes parties labeled as “liberal” or “conservative” obtained in a recent general election of the country. The higher the index, the more conservative the neighborhood tends to be. It can be seen that, after the major protests, more conservative neighborhoods saw a larger decrease in both the number of customers and the median spending, probably due to the fact that people avoided going out for shopping and preferred to save in unstable circumstances. On the days of major events, however, less conservative neighborhoods saw a larger decrease in the number of customers, for the same reason as in the previous analysis that more people from these neighborhoods participated in the protests.

We note that the neighborhoods we analyze here are centrally located, hence it is possible that both the socio-economic status and political preference of the neighborhoods mainly reflect the typical profile of the residents but not that of the visiting customers. The results presented here are therefore mainly behavioral change observed in these neighborhoods. However, according to our data, 54% (46%) transactions in the wealthier (less wealthy) neighborhoods being studied come from their residents or residents of other wealthy (less wealthy) neighborhoods in the city, while 61% (48%) transactions in the conservative (liberal) neighborhoods come from their residents or residents of other conser-



vative (liberal) neighborhoods. Therefore we believe that shopping activities observed in these neighborhoods reflect to a certain extent the behavioral change of people of similar characteristics in terms of wealth and political preference. Tests for statistical significance of the differences between the two groups in Fig. 7 are presented in Table A3(c) in the Appendix.

3.3.3 Gender difference

There may also exist a gender difference in the effect of social events. Figure 8 shows the time series of median change computed over all neighborhoods, for male and female customers.^e We see that, after the two major protests, both male and female customers tended to shop less, although they did not necessarily spend less money. However, demonstrations affected female customers more strongly than their male counterparts on the days of major protests, both in terms of the number of customers going out for shopping (for both days) and the median spending (for Day 77). Tests for statistical significance of the differences are presented in Table A3(d) in the Appendix. Figure A5 in the Appendix further illustrates the differences between four groups of different combinations of gender and political conservatism.

3.3.4 Store category

Finally, Fig. 9 depicts the time series of median change computed over all neighborhoods, for three categories of stores, namely, grocery stores, family clothing stores, and restaurants. Unlike in Fig. 6, Fig. 7 and Fig. 8, the median is weighted by number of stores in each category in the neighborhoods in this case. We see that on days following the major protests, both behavioral indexes decreased for all three categories of stores, except for

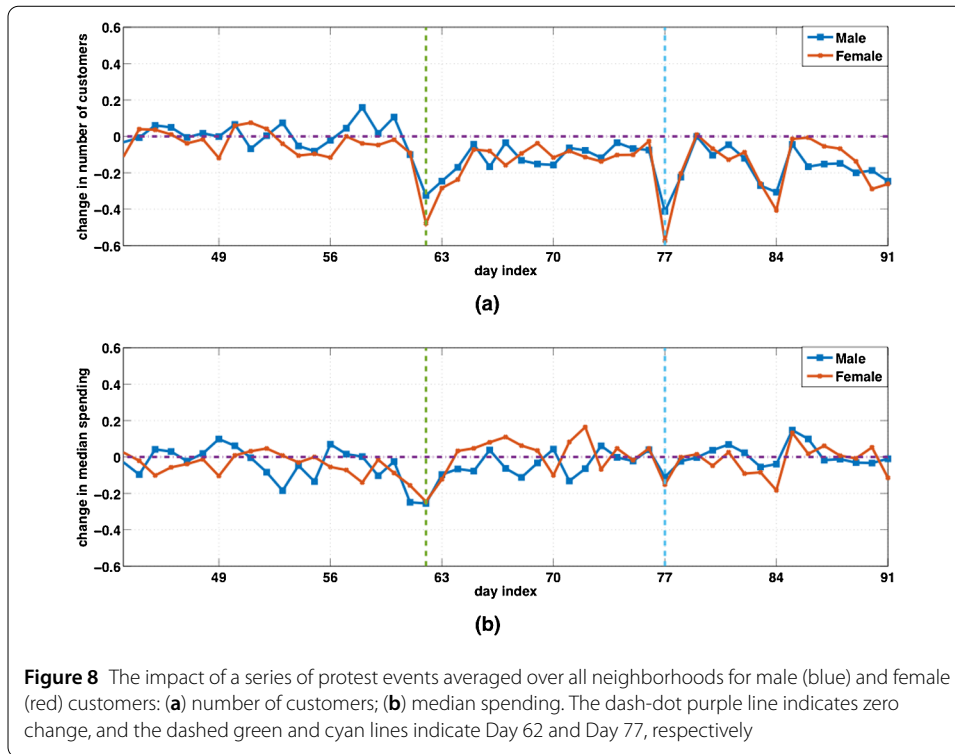


Figure 8 The impact of a series of protest events averaged over all neighborhoods for male (blue) and female (red) customers: **(a)** number of customers; **(b)** median spending. The dash-dot purple line indicates zero change, and the dashed green and cyan lines indicate Day 62 and Day 77, respectively

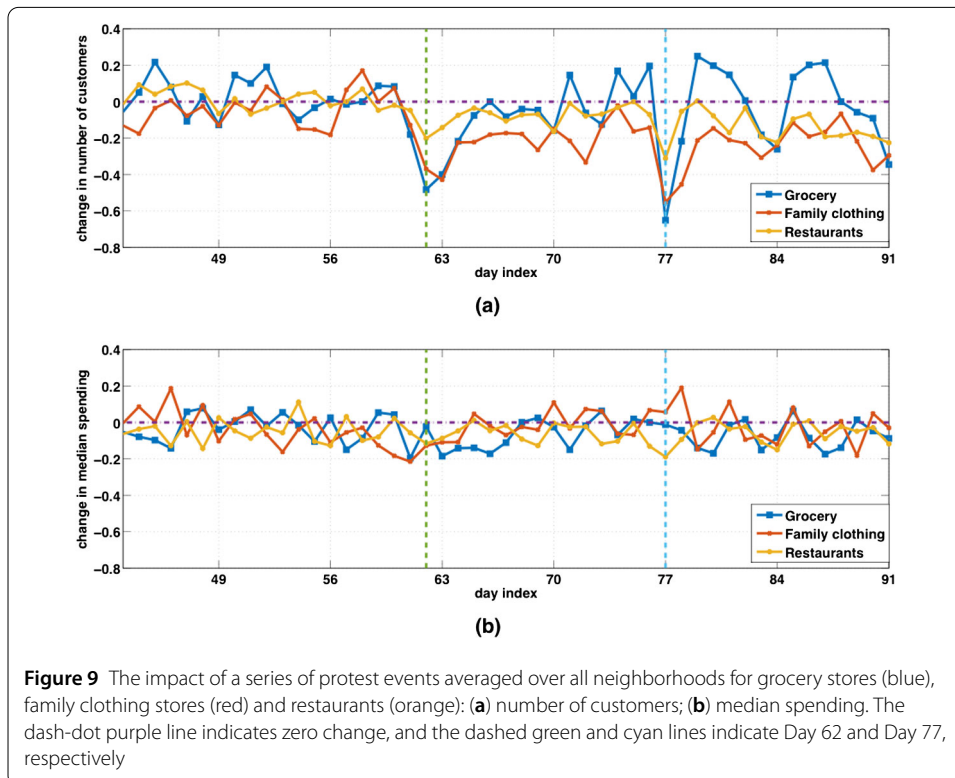


Figure 9 The impact of a series of protest events averaged over all neighborhoods for grocery stores (blue), family clothing stores (red) and restaurants (orange): **(a)** number of customers; **(b)** median spending. The dash-dot purple line indicates zero change, and the dashed green and cyan lines indicate Day 62 and Day 77, respectively

the number of people shopping at grocery stores on mid-week days and median spending at clothing stores. In addition, although there is no clear difference between the three types of stores in terms of the median spending, customers visited family clothing stores less often than restaurants and groceries. Interestingly, on the days of major events, grocery stores showed the largest decrease in terms of customer visits. Tests for statistical significance of the differences are presented in Table A3(e) in the Appendix.

3.4 Integral economic effects

Although societal unrest has led to major changes in people's purchase behavior, particularly on the days of major events, it is interesting to investigate whether there exists an integral economic effect due to displaced or delayed shopping activities.

3.4.1 Displaced shopping

The possibility of displaced shopping may be suggested by Fig. 4(c) where it can be seen that certain neighborhoods further away from the event center, in particular those corresponding to the four squares on the top right, actually saw increased activities on Day 62. In this case, each of the squares represents a single neighborhood (with color indicating the number of stores in the neighborhood), which we analyze in detail as follows.

Increased shopping activities in these four neighborhoods were mainly due to (i) new merchants, and (ii) new customers that did not appear in the reference period. First, the neighborhood corresponding to the red square (100% increase in customers and 60% in spending) saw many payments at a fitness club and, since Day 62 happened to be the first day of the month, these may correspond to people paying subscription fees at the club. As the reference period does not contain first days of the month, this fitness club did not appear as a merchant in the reference period, and therefore the corresponding purchase behavior was not observed before.

Second, the two neighborhoods corresponding to the two light blue squares (220%/120% increase in customers and 20%/100% in spending) saw increased activities due to their close proximity to a new theme park and shopping complex, where those purchases were made on a Saturday (Day 62) at stores that were newly opened in the shopping complex.

Finally, instead of increased activities driven by new merchants, those in the neighborhood corresponding to the last square (60% increase in customers and 100% in spending) were mainly driven by new customers. Indeed, 13 out of the 19 customers purchased in that neighborhood on Day 62 never visited the neighborhood in the reference period, and at the same time they did not visit any neighborhood they used to go on the same day of the week. Therefore, this suggests a pattern of displaced shopping where these customers were shifting their shopping locations, which might be due to the influence of the protest events.

In summary, displaced shopping may have indeed taken place, but being further away from the event center, there could be other factors that influence shopping behavior as well, such as those that came with a special day like Day 62 (being the first day of the month as well as a Saturday that came shortly after the opening of a shopping complex).

3.4.2 Delayed shopping

We see from Fig. 3 that, in terms of both number of customers and median spending, shopping activities generally increased on mid-week days after the weekends when a majority

of the protests took place. For the number of customers, such increase is most obviously in Fig. 9(a) for purchases at grocery stores. Compared to clothing stores and restaurants, grocery stores provide products needed for daily life, and it is therefore possible that people waited during the weekends and got the necessary shopping done shortly after the major protests. On the other hand, increase in median spending on mid-week days were mainly present in neighborhoods within 2 km from the event center, and is most obviously in synchronization with the increased spending in neighborhoods of lower socio-economic status shown in Fig. 6(b). Given that median spending dropped significantly in these neighborhoods on weekends of major protests, this suggests that people may have decided to save in uncertain situations and regained some confidence for make-up purchases after the protest events.

Given the discernible patterns of displaced or delayed shopping activities, however, the integral economic effect of the protest events remains largely negative, as is evident in Fig. A6 where change in total sales is illustrated for neighborhoods that are of different distances from the event center.

4 Discussion

Our study demonstrates the potential of using pervasive and passively collected behavioral data to quantify the impact of localized social shocks and events on an urban population. We found that major events, such as societal unrest due to protests in our case, can alter people's purchase behavior, both in terms of consumers' tendency to visit shops in certain areas and in terms of their willingness to spend money. From a temporal perspective, personal consumption levels recover relatively quickly from the negative influence of the events. Spatially, on days of major events, consumption levels also seem more resilient, although both measures show clear spatial exponential decay. Event effects also differ between groups in the society, defined in terms of demographics, socio-economic status, and political conservatism, and between categories of merchants. Finally, given the results above, the existence of discernible patterns such as displaced or delayed shopping activities suggest that the effect of social events was not always distributed in the way one would imagine, e.g., it is not purely a function of geographical distance, and may help explain the rather unintuitive resilience of (or lack of change in) personal computation as a result of inelastic spending driven by needs.

Our analysis has certain limitations. The data set of credit card transaction records used in this study is based on a sample (about 10%) of all the individual customers of one financial institution and does not include people without credit cards. Therefore, sampling bias could exist and could potentially influence the results. Credit card transaction data may also represent only part of people's daily spendings, as people may choose to pay with cash in certain scenarios. Furthermore, the data set only covers a period of three months, and the observations might be affected by seasonality. Finally, there might exist unobserved external events that bias our results, even though the strong spatiotemporal patterns observed on Days 62 and 77 are probably mainly caused by the major events on those days.

The real-time monitoring of economic behavior can have important social, economic and political implications. As a research tool, it makes it possible to assign quantitative values to the inherently elusive effects of social events. One advantage of the method we present here is the ability to analyze events with relatively high temporal and spatial resolutions. These allow a fine-grained understanding of events, beyond what is often possible

with traditional measures. Also, the measures are less affected by demand characteristics or other biases, inherent in methods such as surveys. Such understanding and measures can potentially be used to develop timely and effective policies targeted at specific communities and demographic groups, and can often be critical in emergency situations.

With our method we can also evaluate statements made, for instance, by politicians or the media, such as that people are greatly upset by certain events. We can compare them to the observed changes in behavior, as expressed in the number of customers doing purchases and the average amount spent in a purchase. If people go about their activities as usual, the events probably have less impact, compared to when measures of behavior differ strongly from the usual values. At a practical level, these analyses may allow authorities to prepare for timely interventions to minimize possible negative consequences, possibly through a prediction mechanism that could estimate the recovery time of individual economic activities at an early stage of the event.

The framework we present here complements, rather than replaces, other methods such as traditional economic indicators or surveys. Combinations of it with other tools help create a comprehensive picture of the dynamic response of a population to social events.

Appendix

We present in Table A1 and Table A2 the following statistics about purchase activity comparison on the reference days and event days (Day 62 and Day 77), for each neighborhood within 4 km from the event center:

1. the minimum/maximum number of customers on the reference days for the event day;
2. the decrease in number of customers on Day 62 compared to the average of its reference days;
3. the minimum/maximum median spending amount on the reference days for the event day;
4. the decrease in median spending amount on Day 62 compared to the average of its reference days;
5. the minimum/maximum number of transactions on the reference days for the event day;
6. the decrease in number of transactions on Day 62 compared to the average of its reference days;
7. the minimum/maximum total sales amount on the reference days for the event day;
8. the decrease in total sales amount on Day 62 compared to the average of its reference days;

We test the statistical significance of differences between the different curves shown in Fig. 3(a), (b), Fig. 6(a), (b), Fig. 7(a), (b), Fig. 8(a), (b) and Fig. 9(a), (b). Our tests focus on the following different periods:

1. The overall period from Day 43 to Day 91;
2. The days before the first major protest began, namely, from Day 43 to Day 61;
3. The days following the two major protests, namely, from Day 62 to Day 91, but excluding Days 62 and 77;
4. On Day 62;
5. On Day 77.

Table A1 Statistics about purchase activities on both reference days and Day 62, for each neighborhood within 4 km from the event center

Distance from event center (km)	Number of customers			Median spending amount			Number of transactions			Total sales amount		
	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62
0.22	13.00	26.00	18.00	15.00	40.50	-22.24	13.00	26.00	18.60	366.25	1526.46	849.37
0.33	12.00	26.00	7.80	15.50	63.00	25.12	13.00	28.00	6.20	655.50	1658.97	667.03
0.36	60.00	105.00	78.80	39.99	52.50	29.13	65.00	120.00	91.20	4429.38	12,782.01	7428.72
0.44	23.00	33.00	3.40	38.19	50.90	19.68	25.00	35.00	0.20	1376.15	2219.74	676.15
0.66	88.00	114.00	89.20	41.50	53.90	29.68	94.00	126.00	98.60	7533.84	11,605.94	9675.07
1.01	196.00	218.00	127.20	73.18	82.00	19.12	241.00	276.00	166.60	37,109.77	53,761.52	36,545.05
1.13	95.00	150.00	102.20	33.00	56.00	20.35	99.00	197.00	121.80	7475.87	15,507.40	8961.38
1.18	23.00	29.00	6.40	37.10	76.50	20.10	25.00	30.00	4.80	1993.55	4357.06	2124.65
1.45	9.00	18.00	-0.20	14.47	30.70	-12.76	10.00	22.00	0.80	175.56	1452.04	-1443.47
1.52	79.00	107.00	75.40	39.80	49.50	21.06	82.00	113.00	78.00	6962.00	19,962.29	8902.90
1.56	19.00	30.00	10.40	50.00	72.50	-14.20	19.00	33.00	11.60	1407.76	2924.73	-931.71
1.64	49.00	67.00	43.00	49.92	63.00	22.19	55.00	76.00	48.00	5096.61	7225.35	5270.46
1.73	118.00	135.00	53.40	29.90	36.86	-1.19	131.00	159.00	63.20	7375.94	9452.52	2284.04
1.80	26.00	42.00	6.40	26.84	45.06	8.83	30.00	47.00	10.20	1275.25	3682.94	1361.21
1.88	35.00	49.00	11.00	21.24	28.25	-12.46	40.00	51.00	10.00	1565.38	2079.06	335.96
1.91	173.00	197.00	76.40	48.56	59.00	11.79	198.00	229.00	97.40	19,375.87	28,661.25	14,788.83
2.06	56.00	82.00	26.60	29.97	41.62	7.23	59.00	88.00	22.60	4026.21	5333.42	2597.16
2.07	15.00	26.00	-1.20	21.76	39.08	-1.65	17.00	26.00	-1.80	790.73	1834.64	-274.68
2.11	11.00	26.00	-2.40	20.00	40.00	5.99	11.00	29.00	-2.00	513.56	1380.94	-213.42
2.18	20.00	41.00	15.00	13.75	19.25	-1.05	20.00	43.00	16.60	459.00	736.69	314.23
2.59	295.00	384.00	138.20	49.99	57.10	9.77	389.00	552.00	180.40	44,508.80	60,045.30	21,319.15
2.60	34.00	46.00	-2.00	17.80	34.98	-5.31	37.00	54.00	-1.00	1475.96	6773.80	-512.23
2.64	15.00	28.00	7.60	35.00	70.84	-42.25	16.00	28.00	8.20	1471.68	9173.52	2314.25
2.70	12.00	15.00	-0.80	13.75	25.50	-48.48	12.00	17.00	-0.80	531.59	1276.54	-673.34
2.85	12.00	19.00	-1.80	17.85	25.00	2.78	12.00	24.00	-2.00	255.51	776.19	21.78
2.87	21.00	40.00	2.60	18.37	32.99	3.88	23.00	44.00	3.80	1012.07	2488.92	204.27
2.97	68.00	84.00	11.00	30.40	47.45	10.53	75.00	98.00	11.60	4350.82	7438.82	2034.89
3.10	55.00	78.00	1.80	23.94	43.80	11.34	66.00	85.00	3.80	3019.73	3915.53	1145.75

Table A1 (Continued)

Distance from event center (km)	Number of customers			Median spending amount			Number of transactions			Total sales amount		
	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62
3.15	5.00	17.00	-1.80	12.15	34.22	-2.85	5.00	18.00	-3.20	56.31	591.19	-113.85
3.29	13.00	23.00	3.40	14.38	41.33	6.21	13.00	24.00	2.60	458.44	1152.67	-353.15
3.29	27.00	43.00	-7.00	18.28	39.27	4.06	28.00	50.00	-5.60	1338.60	2590.01	-331.94
3.35	66.00	84.00	11.00	37.61	51.00	0.99	70.00	88.00	7.20	5268.75	9210.19	15.55
3.44	11.00	18.00	0.40	21.05	32.60	-12.26	11.00	20.00	0.80	476.67	875.30	76.61
3.54	11.00	23.00	13.20	27.35	81.03	17.22	12.00	24.00	14.00	702.17	6732.62	2195.74
3.78	163.00	205.00	40.60	26.69	39.85	4.52	193.00	229.00	56.60	11,748.51	23,198.06	9441.16
3.85	25.00	35.00	0.80	21.91	31.43	1.63	26.00	38.00	-5.20	725.72	3774.05	130.77
3.87	6.00	18.00	-0.20	10.15	30.69	-3.27	6.00	18.00	-2.00	154.68	4480.16	761.56
3.92	29.00	51.00	13.40	22.99	43.00	-7.88	34.00	63.00	13.60	1715.98	5400.22	1925.14
3.97	7.00	13.00	-0.80	13.44	44.10	4.33	7.00	16.00	0.20	189.52	666.88	30.80
3.99	65.00	107.00	31.00	25.00	46.80	0.95	68.00	114.00	33.00	4636.95	13,369.05	3718.05

Table A2 Statistics about purchase activities on both reference days and Day 77, for each neighborhood within 4 km from the event center

Distance from event center (km)	Number of customers			Median spending amount			Number of transactions			Total sales amount		
	Min of ref. days	Max of ref. days	Decrease on Day 77	Min of ref. days	Max of ref. days	Decrease on Day 77	Min of ref. days	Max of ref. days	Decrease on Day 77	Min of ref. days	Max of ref. days	Decrease on Day 77
0.22	900	1600	11.50	14.00	30.00	-63.33	900	17.00	12.17	260.63	714.82	266.26
0.33	800	1600	11.33	15.88	57.50	-34.46	800	22.00	13.00	332.00	1188.90	663.03
0.36	4200	5900	50.33	26.50	59.97	23.04	4800	76.00	60.67	3976.93	8816.86	6196.56
0.44	1900	2800	9.00	28.92	42.38	-12.26	2100	31.00	10.17	1071.89	1351.47	428.51
0.66	4900	8300	61.67	39.99	59.99	32.01	5100	89.00	66.17	3864.26	9830.46	5578.12
1.01	10300	12300	77.33	49.50	79.97	12.62	12100	151.00	91.17	14242.68	20257.10	13401.79
1.13	4500	10000	69.83	30.00	44.00	-6.10	5900	123.00	82.67	41206.0	8067.87	6212.16
1.18	1400	2300	7.83	17.56	37.58	0.95	1400	23.00	9.33	664.42	2637.18	1374.48
1.45	700	1800	1.00	7.70	32.50	6.33	900	21.00	2.33	242.89	3805.41	673.86
1.52	4800	7400	49.33	30.00	50.00	17.23	5100	79.00	53.33	3022.43	8592.50	5918.30
1.56	700	2000	7.83	29.25	141.50	-9.38	800	22.00	8.67	777.16	1375.35	628.29
1.64	900	3800	23.17	33.69	48.50	-81.01	900	42.00	24.17	396.83	3717.19	1607.31
1.73	8100	12000	32.33	28.84	37.36	0.28	9100	142.00	39.50	3570.15	7902.30	2611.83
1.80	1900	3100	-4.17	22.70	33.82	1.29	1900	34.00	-3.17	627.89	1594.24	-63.66
1.88	2900	4400	26.83	16.47	29.50	4.38	3500	50.00	31.50	993.79	1950.08	1310.44
1.91	10400	15400	61.00	39.96	57.79	3.45	12300	180.00	70.67	10301.30	18156.04	8309.15
2.06	3900	5300	11.17	25.00	41.02	-3.91	4100	57.00	11.33	1820.16	3737.76	492.95
2.07	1500	2800	9.83	26.35	50.12	8.60	1600	29.00	9.67	717.11	1779.31	703.60
2.11	900	2100	2.33	21.25	44.00	-13.11	900	23.00	2.83	297.31	1988.85	80.12
2.18	1300	2600	0.67	11.75	20.00	0.46	1500	29.00	0.33	306.63	658.75	53.35
2.59	26000	35900	157.83	48.67	69.22	7.56	36100	486.00	227.83	37275.03	56893.04	18664.75
2.60	3200	4500	5.00	22.26	30.00	-3.25	3500	50.00	4.83	1446.06	4502.02	381.34
2.64	300	900	3.50	48.96	179.40	54.31	300	9.00	3.83	154.97	1589.10	601.25
2.70	700	1500	7.17	22.10	37.80	7.81	700	15.00	7.33	324.99	3714.85	1228.37
2.85	1100	2300	4.33	14.47	28.23	2.13	1400	29.00	6.67	312.47	917.88	212.82
2.87	1900	3700	3.83	28.62	38.02	9.82	1900	38.00	2.83	599.71	3470.42	616.43
2.97	5300	7400	13.83	30.20	49.65	12.39	5800	78.00	17.33	3162.86	5199.08	2357.99
3.10	3000	5500	-2.83	20.55	27.79	6.85	3200	59.00	-4.00	1099.56	2495.46	-163.64

Table A2 (Continued)

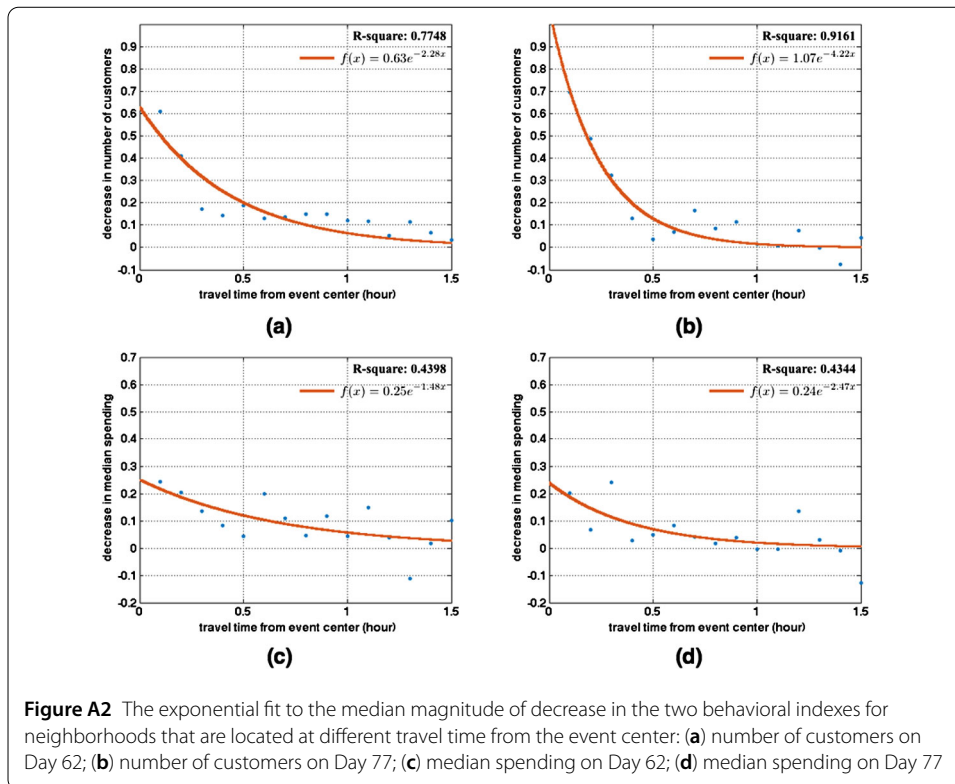
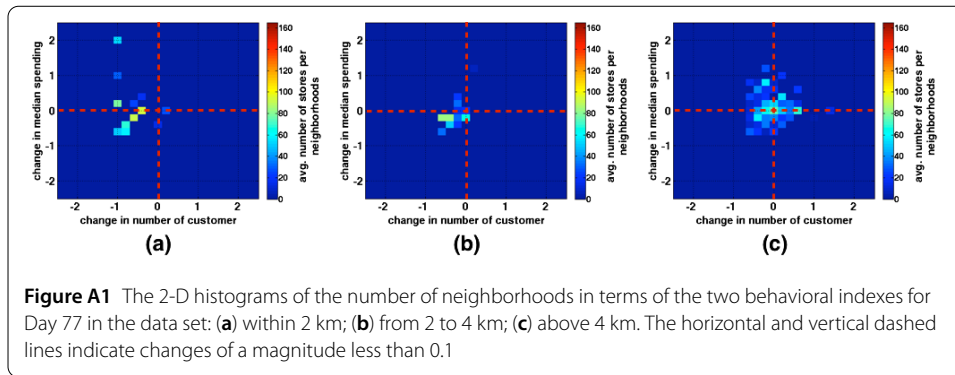
Distance from event center (km)	Number of customers			Median spending amount			Number of transactions			Total sales amount		
	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62	Min of ref. days	Max of ref. days	Decrease on Day 62
3.15	8.00	16.00	3.17	14.50	35.64	9.96	9.00	18.00	-0.17	217.18	569.31	69.51
3.29	12.00	21.00	2.83	18.74	58.70	1.32	14.00	23.00	3.83	468.35	1209.39	125.60
3.29	27.00	36.00	3.67	24.40	40.00	-6.04	27.00	41.00	5.83	1039.82	1838.75	52.65
3.35	73.00	84.00	28.50	44.13	61.64	15.03	82.00	90.00	24.50	6587.63	7784.76	3232.86
3.44	11.00	25.00	4.67	16.27	35.70	-2.06	12.00	28.00	5.50	300.32	1219.77	316.85
3.54	13.00	18.00	1.67	27.93	64.62	0.63	14.00	19.00	1.67	1043.04	2098.67	-213.02
3.78	150.00	252.00	60.83	24.15	32.01	6.59	171.00	272.00	65.50	8108.18	17,572.56	4656.21
3.85	24.00	34.00	3.33	24.45	32.69	-2.05	25.00	37.00	2.67	983.01	2442.08	427.52
3.87	7.00	18.00	3.67	14.75	37.00	-7.73	8.00	20.00	1.67	180.77	909.34	81.45
3.92	27.00	42.00	11.50	20.78	33.72	10.75	32.00	55.00	16.17	1408.99	5795.05	1834.46
3.97	8.00	14.00	-1.00	10.75	33.10	-24.68	8.00	16.00	-1.50	143.59	534.13	-250.21
3.99	43.00	58.00	21.83	30.98	53.84	11.60	46.00	62.00	21.83	2464.06	8247.45	1905.82

Table A3 (Continued)

Time period	(c) Political conservatism				
	Compared with zero (for number of customers)				
	Conservative	Liberal	Diff. btw. conservative and liberal	Conservative	Liberal
	Compared with zero (for median spending)			Conservative	Liberal
					Diff. btw. conservative and liberal
Overall	<10e-4	<10e-4	0.142	0.007	0.595
Before first major protests	0.469	0.601	0.904	0.421	0.601
Following major protests	<10e-4	<10e-4	0.011	0.020	0.087
Day 62	<10e-4	<10e-4	<10e-4	<10e-4	<10e-4
Day 77	<10e-4	<10e-4	<10e-4	<10e-4	<10e-4
Time period	(d) Gender				
	Compared with zero (for number of customers)				
	Male	Female	Diff. btw. male and female	Male	Female
	Compared with zero (for median spending)			Male	Female
					Diff. btw. male and female
Overall	<10e-4	<10e-4	0.084	0.040	0.162
Before first major protests	0.573	0.071	0.020	0.198	0.011
Following major protests	<10e-4	<10e-4	0.767	0.274	0.601
Day 62	<10e-4	<10e-4	<10e-4	<10e-4	<10e-4
Day 77	<10e-4	<10e-4	<10e-4	<10e-4	<10e-4

Table A3 (Continued)

Time period	(e) Store category									
	Compared with zero (for number of customers)					Compared with zero (for median spending)				
	Grocery	Clothing	Restaurants	Diff. btw. grocery and clothing	Diff. btw. grocery and restaurants	Grocery	Clothing	Restaurants	Diff. btw. grocery and clothing	Diff. btw. grocery and restaurants
Overall	0.559	<10e-4	<10e-4	<10e-4	0.201	<10e-4	0.040	<10e-4	0.129	0.870
Before first major protests	0.500	0.085	0.309	0.024	0.968	0.070	0.117	0.009	0.687	0.546
Following major protests	0.534	<10e-4	<10e-4	<10e-4	0.020	<10e-4	0.219	<10e-4	0.015	0.412
Day 62	<10e-4	<10e-4	<10e-4	significant	significant	significant	<10e-4	<10e-4	non-significant	non-significant
Day 77	<10e-4	<10e-4	<10e-4	significant	significant	significant	0.004	<10e-4	significant	significant



We use non-parametric statistical tests that do not require assumption of a probability distribution, e.g., a normal distribution, of the data. Specifically, for (1) to (3), we use a Wilcoxon signed-rank test to test the statistical significance of difference between pairs of different time series, as well as difference between the values of time series of each group and zeros. For (4) to (5), we construct samples for each group that consist of the index of each neighborhood in the group repeated by the number of stores in that neighborhood. We then use a Wilcoxon signed-rank test to test the statistical significance of whether each group has a median different from zero. For pairwise comparison between the groups, we use a Wilcoxon rank-sum test for 2-category comparisons (socio-economic status, political conservatism, and gender), and a Kruskal–Wallis test followed up with post hoc comparisons using the Dunn–Sidak method [25] for 3-category comparisons (distance and store category). In Table A3 we report the p -value of the statistical tests, where bold font indicates that a null hypothesis of no difference can be rejected at the 5% significance

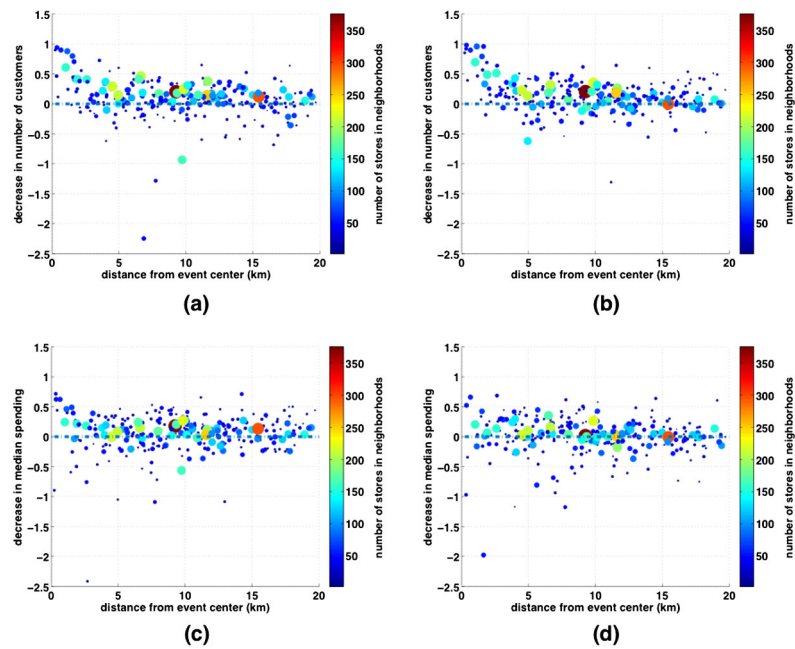


Figure A3 The decrease in the two behavioral indexes for neighborhoods that are located at different distances from the event center. (a) number of customers on Day 62; (b) number of customers on Day 77; (c) median spending on Day 62; (d) median spending on Day 77. Both the size and color of the circles indicate the number of stores in each neighborhood

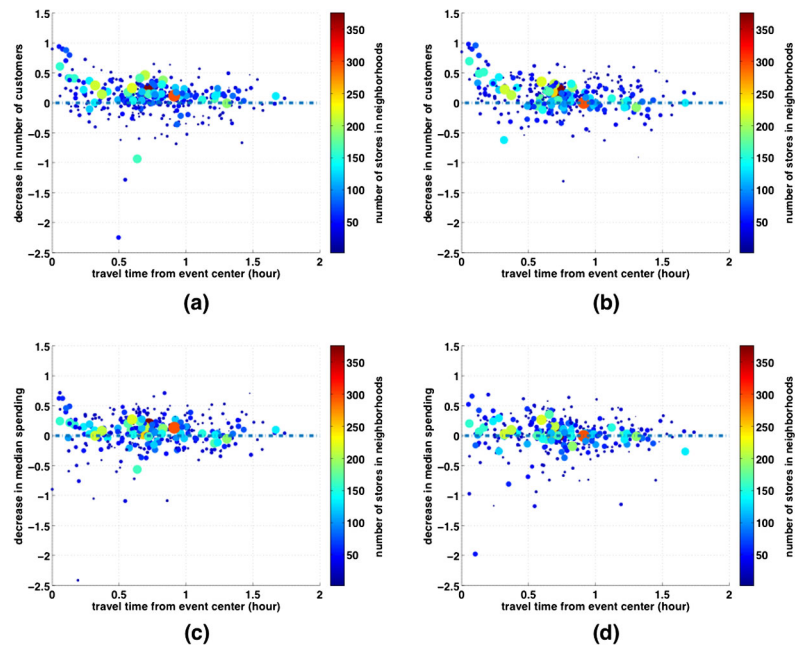
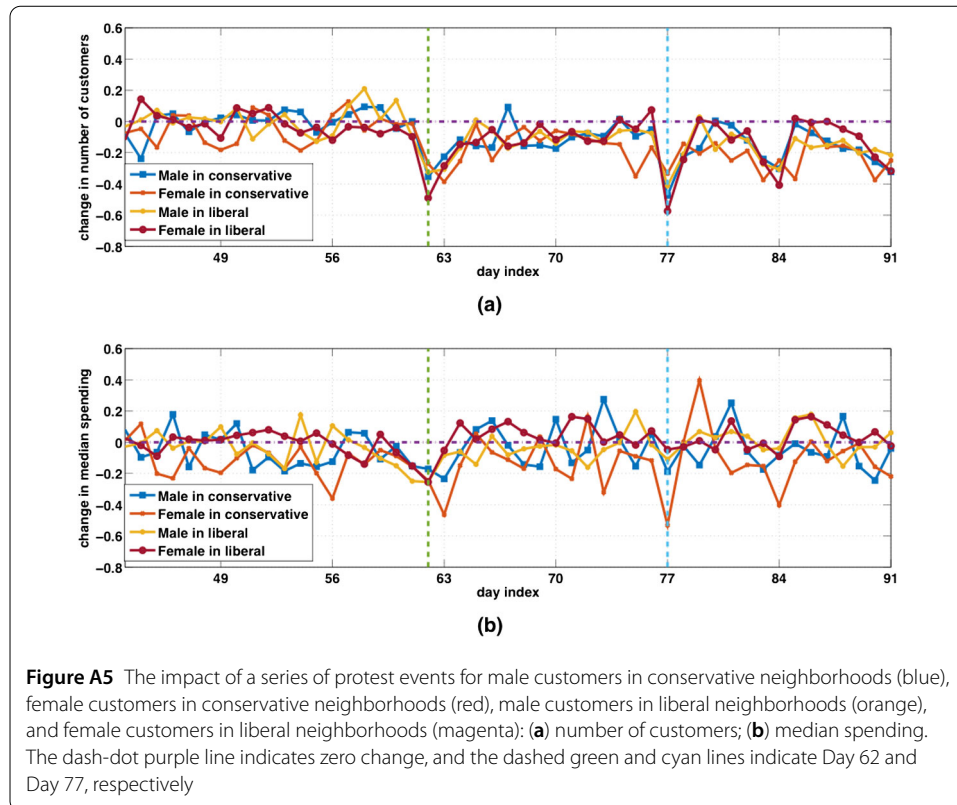


Figure A4 The decrease in the two behavioral indexes for neighborhoods that are located at different travel time from the event center. (a) number of customers on Day 62; (b) number of customers on Day 77; (c) median spending on Day 62; (d) median spending on Day 77. Both the size and color of the circles indicate the number of stores in each neighborhood



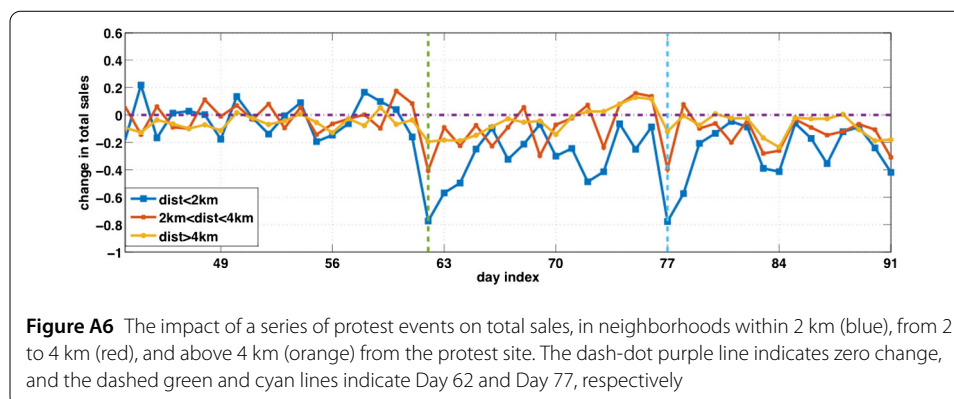
level. Notice that for 3-category comparisons we report the statistical significance directly because there are no p -value for the pairwise comparison.

Figure A1 show the 2-D histograms that illustrate the combined effect of the events on the two independent behavioral indexes for Day 77. It can be seen that a vast majority of the neighborhoods within 4 km of the event center saw less customers than usual, even though some of them actually attracted more spending. On the other hand, such effect was not obvious for neighborhoods that are more than 4 km further away.

Figure A2 shows the event effect as a function of the estimated travel time (by car) from the event center, thus taking into account geographical constraint and transportation accessibility. As we can see, the exponentially decaying patterns are similar to those in Fig. 5.

In addition to the average results in Fig. 5 and Fig. A2, Fig. A3 and Fig. A4 show the event effect on each single individual neighborhoods, where both the size and color of the circles indicate the number of stores in each neighborhood. We see that neighborhoods with large number of stores generally saw reduced amount of shopping activity in terms of both the number of customers and the spending amount, while such reduction was more significant for the smaller neighborhoods that are closer to the event center. This is consistent with the exponentially decaying patterns we have seen in Fig. 5 and Fig. A2.

Figure A5 illustrates the differences between four groups based on combinations of gender and political conservatism, namely, male customers in conservative neighborhoods (blue), female customers in conservative neighborhoods (red), male customers in liberal neighborhoods (orange), and female customers in liberal neighborhoods (magenta). We see that, on both days of major protests, liberal females reduced their shopping activi-



ties most, while conservative females tended to save money more than other groups (for Day 77). The latter was also generally observed for days following the major protests.

Figure A6 shows the effect of the events on total sales, in neighborhoods within 2 km (blue), from 2 to 4 km (red), and above 4 km (orange) from the protest site. We see that, although displaced or delayed shopping activities may have taken place, the integral economic effect of the protest events remains largely negative.

Additional material

Additional file 1: Dynamic change in the number of customers and median spending from Day 43 to Day 91. The four rings correspond to the neighborhood that contains the event center as well as the neighborhoods that are within 2 km, between 2 to 4 km away, and more than 4 km away from the center. (MOV 3.4 MB)

Acknowledgements

X. Dong was supported by a Swiss National Science Foundation Mobility fellowship while completing this work. The authors are grateful to the financial institution that provided the credit card transaction data for this research.

Availability of data and materials

Data sufficient to reproduce all results in this paper will be made available upon request.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XD and JM conceived and designed the study. XD, JM, ES and BB conducted the analyses. XD, JM and ES wrote the manuscript. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Author details

¹MIT Media Lab, Cambridge, USA. ²Department of Industrial Engineering, Tel Aviv University, Tel Aviv, Israel. ³Sabancı School of Management, Sabancı University, Tuzla/Istanbul, Turkey.

Endnotes

- ^a The two latter steps aim to remove inactive customers and stores to prevent them from biasing subsequent analyses. The specific thresholds of ten customers and ten stores we used here have little effect on the amount of filtered data: with alternative thresholds of five customers/stores or three, 2.0 or 2.1 million transaction records are left after the filtering process.
- ^b An alternative would be to compute a z-score that also captures the variability of the static. However, computing sample standard deviation over a relatively small number of points may lead to large variation in the z-score hence biases the analysis afterwards. We therefore choose the relative deviation from the mean, which is a more stable measure and is also adopted in [20, 22].
- ^c Except that we remove three holidays in these six weeks when shopping activities were clearly boosted and would introduce obvious bias in the analysis.
- ^d Notice that the behavioral indexes are computed as relative changes with respect to a reference period, therefore the factor that activities may decrease further away from the city center due to less shops available has already been taken into account.

- ^e In our data set, we have 34% female customers who account for 38% of the transactions, which is moderately imbalanced. However, since the behavioral indexes are computed at the neighborhood level as relative changes with respect to a reference period, and gender split is assumed to stay reasonably stable across the reference and analysis periods, such imbalance would not cause an issue in the analysis and statistical tests presented in the paper.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 23 July 2017 Accepted: 18 March 2018 Published online: 13 April 2018

References

- Long PT, Perdue RR (1990) The economic impact of rural festivals and special events: assessing the spatial distribution of expenditures. *J Travel Res* 28(4):10–14
- Burgan B, Mules T (1992) Economic impact of sporting events. *Ann Tour Res* 19(4):700–710
- Crompton JL, McKay SL (1994) Measuring the economic impact of festivals and events: some myths, misapplications and ethical dilemmas. *Festiv Manage Event Tour* 2:33–43
- Mondello MJ, Rishe P (2004) Comparative economic impact analyses: differences across cities, events, and demographics. *Econ Dev Q* 18(4):331–342
- Getz D (2005) Event management and event tourism, 2nd edn. Cognizant Communication Corporation, New York
- Herrero LC, Sanz JA, Devesa M, Bedate A, del Barrio MJ (2006) The economic impact of cultural events: a case-study of Salamanca 2002, European capital of culture. *Eur Urban Reg Stud* 13(1):41–57
- Jong-A-Pin R (2009) On the measurement of political instability and its impact on economic growth. *Eur J Polit Econ* 25(1):15–29
- Personal income and outlays, U.S. Department of Commerce, Bureau of Economic Analysis (BEA). <http://www.bea.gov/newsreleases/national/pi/pinewsrelease.htm>
- Consumer confidence index, The Conference Board. <https://www.conference-board.org/data/consumerconfidence.cfm>
- Surveys of consumers, University of Michigan. <http://www.sca.isr.umich.edu>
- Carroll CD, Fuhrer JC, Wilcox DW (1994) Does consumer sentiment forecast household spending? If so, why? *Am Econ Rev* 84:1397–1408
- Lazer D, Pentland A, Adamic L, Aral S, Barabási A-L, Brewer D, Christakis N, Contractor N, Fowler J, Gutmann M, Jebara T, King G, Macy M, Roy D, Alstyn MV (2009) Computational social science. *Science* 323(5915):721–723
- Pentland A (2014) Social physics: how good ideas spread. Penguin, Baltimore
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the international conference on World Wide Web, pp 851–860
- Yates D, Paquette S (2011) Emergency knowledge management and social media technologies: a case study of the 2010 Haitian earthquake. *Int J Inf Manag* 31(1):6–13
- Yin J, Lampert A, Cameron M, Robinson B, Power R (2012) Using social media to enhance emergency situation awareness. *IEEE Intell Syst* 27(6):52–59
- Imran M, Castillo C, Diaz F, Vieweg S (2015) Processing social media messages in mass emergency: a survey. *ACM Comput Surv* 47(4):67–16738
- Martínez EA, Rubio MH, Martínez RM, Arias JM, Patane D, Zerbe A, Kirkpatrick R, Luengo-Oroz M (2016) Measuring economic resilience to natural disasters with big economic transaction data. In: Proceedings of the data for good exchange
- Kapoor A, Eagle N, Horvitz E (2010) People, quakes, and communications: inferences from call dynamics about a seismic event and its influences on a population. In: Proceedings of AAAI symposium on artificial intelligence for development, pp 51–56
- Bagrow JP, Wang D, Barabási A-L (2011) Collective response of human populations to large-scale emergencies. *PLoS ONE* 6(3):17680
- Altshuler Y, Fire M, Shmueli E, Elovici Y, Bruckstein A, Pentland A, Lazer D (2013) The social amplifier—reaction of human communities to emergencies. *J Stat Phys* 152(3):399–418
- Gao L, Song C, Gao Z, Barabási A-L, Bagrow JP, Wang D (2014) Quantifying information flow during emergencies. *Sci Rep* 4:3997
- Einav L, Levin J (2014) Economics in the age of big data. *Science* 346(6210):715
- Dong X, Jahani E, Morales AJ, Bozkaya B, Lepri B, Pentland A (2016) Purchase patterns, socioeconomic status, and political inclination. In: International conference on computational social science
- Hochberg Y, Tamhane AC (1987) Multiple comparison procedures. Wiley, New York