Does Information Provision Spur Follow-on Contributions in Online Communities?

Evidence from Crowdsourced Mapping

Abhishek Nagaraj
NBER (2016-17) / UC Berkeley, Haas (2017–)
nagaraj@berkeley.edu — +1 (617) 401 1662

BerkeleyHaas
Haas School of Business
University of California Berkeley

Abstract

Does the provision of information spur follow-on knowledge contributions in online communities? I analyze this question within the context of OpenStreetMap, a Wikipedia-style community-contributed mapping project used widely on the internet. I exploit a natural experiment during the initial development of OpenStreetMap in the US, where about 60% of the 3100 counties in the United States benefited from the provision of a relatively complete baseline map upon which the community could build, while the other 40% accidentally received a largely incomplete government map. By tracking data on about 350 million contributions made by about 577,000 contributors across both kinds of counties, I estimate the causal effects of information provision on follow-contributions in online communities. I find that the provision of basic knowledge hurts the development of online communities in affected counties, by reducing the number of follow-on contributions as well as the number of active contributors. TIGER-affected regions also seem to have lower levels of eventual quality as compared to control counties. TIGER information seems to discourage contributors from adding follow-on information by crowding out their ability to develop ownership over the underlying information.

Research Question

Why do some online communities become large and vibrant over time, while most do not grow beyond a few users? This paper investigates the impact of seeding communities with information in their early stages on their long-term development. In particular, I focus on the role of the provision of basic information on influencing the amount of follow-on contributions and contributor activity in online communities.

What is OpenStreetMap

Figure 1: A group of OpenStreetMap users on a "mapping party"

- Online community to build a "map of the world"
- About 3.2 million registered users
- Inspired by Wikipedia, allows anyone to add, edit or delete objects from the map

OpenStreetMap TIGER Experiment

Figure 2: A map showing the distribution of affected counties

During the initial development on OpenStreetMap in the US, community members decided to seed the map with data from the US Census TIGER project. The TIGER data itself was being updated due to which about 60% of counties in the US had relatively complete and accurate data while the other 40% contained largely missing and incomplete data. This accident introduced unintentional variation in the quality of the initial information on OpenStreetMap, which is the natural experiment at the heart of this paper.

Comparing TIGER and Control Counties

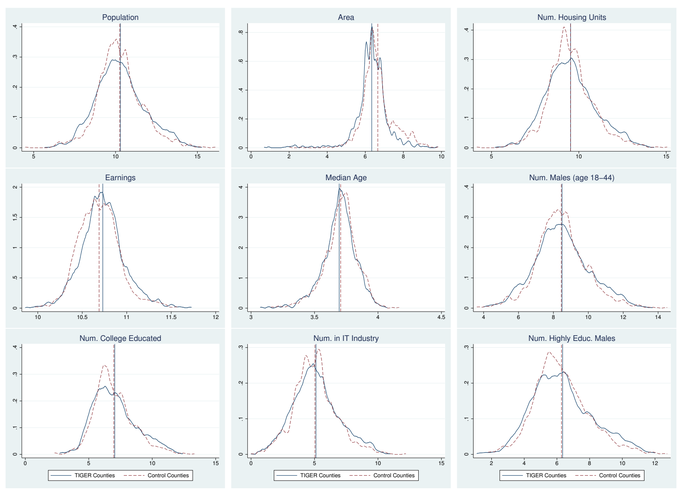

Figure 3: Comparing covariates between TIGER and Control counties in terms of demographic, geographic and economic variables

Example of Quasi-Experimental Variation

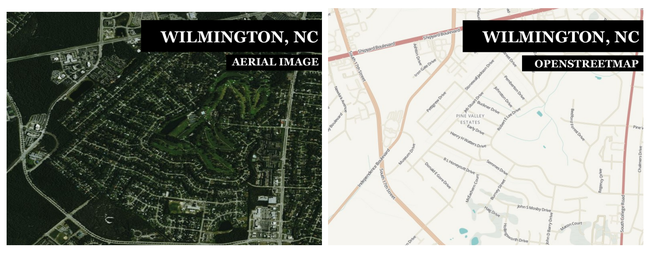

Figure 4: TIGER counties benefit from an initial endowment of basic information

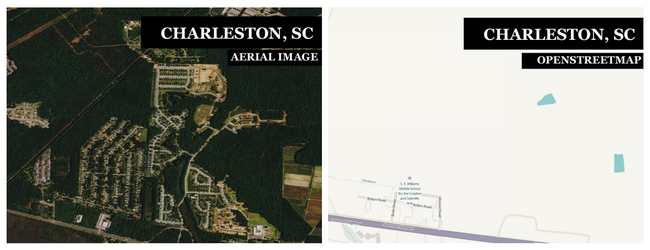

Figure 5: Control counties have missing and incomplete initial information

Real Implications of the TIGER Experiment

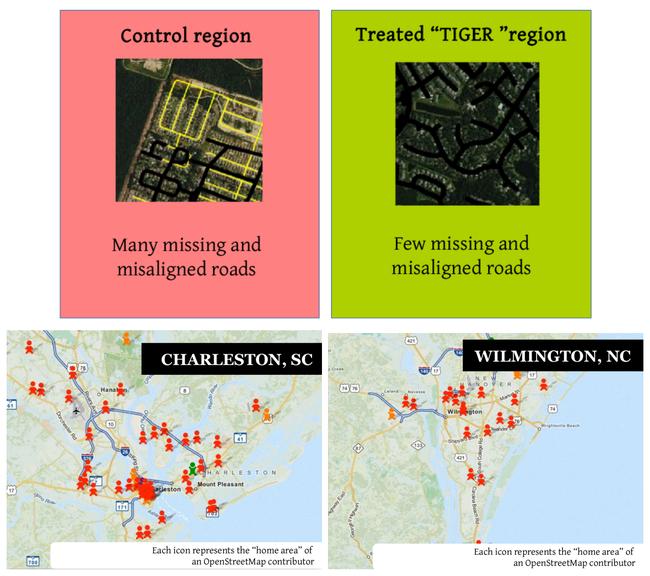

Figure 6: Contributors in Charleston (left) vs Wilmongton(right)

Empirical Results

I analyze the impact of the TIGER project on OpenStreetMap activity using the county-quarter balanced panel. These data measure three key outcome variables at the county-quarter level for the mainland US: contributions and users. *Follow-on Contributions* measures the number of follow-on contributions to the map, *Users* measures the number of users making contributions and *New Users* measures number of first-time users making contributions. I then estimate regressions of the following form using the county-quarter panel:

$$Ln(Y_{it}) = \alpha + \beta_1 \times Post_t \times Treat_i + \gamma_i + \delta_t + \epsilon_{it}$$

where γ_i and δ_t represent county and year fixed effects respectively for county i in quarter t. $Post_t$ equals one for all quarters after December 2007, and $Treat_t$ equals one for all TIGER counties.

Baseline Results

Table 1: Impact of TIGER on OpenStreetMap

	Ln(Follow-on)		Ln(Users)		Ln(New-Users)	
	(1)	(2)	(3)	(4)	(5)	(6)
Post X Treat	-0.109**	-0.114**	-0.0365*	-0.0449***	-0.0366**	-0.0401***
	(0.0513)	(0.0484)	(0.0188)	(0.0170)	(0.0169)	(0.0155)
County FE	No	Yes	No	Yes	No	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes	Yes
N	121212	121212	121212	121212	121212	121212

Figure 7: Time-varying Impacts of TIGER on Follow-on Contributions

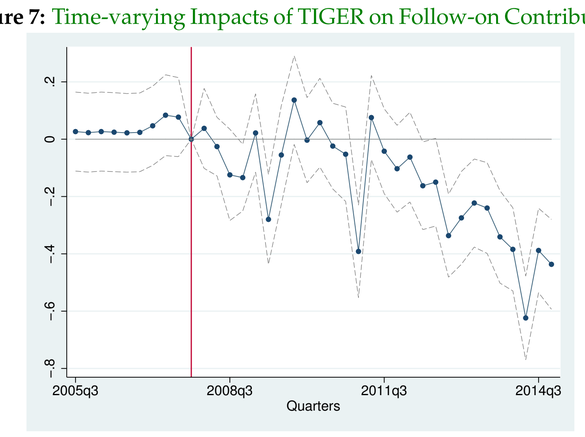

Table 2: Comparing Eventual Quality Between TIGER and Control Counties

	Log(Quality)	Log(Quality)	Log(Quality)
TIGER	0.104	0.121*	0.138**
	(0.0651)	(0.0631)	(0.0631)
Distance	0.0364***	0.0373***	0.0371***
	(0.00221)	(0.00221)	(0.00219)
Area/Pop Controls	Yes	Yes	Yes
Pop. Density FE	No	Yes	Yes
Income/Demo. Controls	No	No	Yes
N	99136	99136	99136

Testing the Ownership Mechanism

Hypothesized mechanism for the negative effect of information on follow-on contributions: Contributors are more likely to build on existing knowledge when they develop "ownership" over the basic knowledge, and seeding information crowds out the ability of contributors to develop such ownership. Direct implications of this theory are tested. Impact of TIGER should be diminished for (1) un-related follow-on knowledge (buildings vs. highway tags), (2) new-users as compared to existing users and (3) in rapidly-changing urban areas.

Table 3: Testing the 'Ownership' Mechanism

	Distant vs Near		Users		Urban vs. Rural
	Distant	Incremental	New-Users	Old-Users	Follow-On
Post X Treat	0.165***	-0.196***	-0.0401***	-0.0236	-0.314***
	(0.0444)	(0.0526)	(0.0155)	(0.0185)	(0.0668)
Post X Treat X Urban					0.382***
					(0.0964)
County FE	Yes	Yes	Yes	Yes	Yes
Quarter FE	Yes	Yes	Yes	Yes	Yes
N	121212	121212	121212	121212	121212

Conclusions

Even though many platforms often "seed" online communities with initial information, this study finds that such practices might be misguided. In OpenStreetMap, the provision of TIGER information reduced the likelihood of follow-on contributions from the community in affected counties, and lowered user activity as well as overall map quality in the long-run. While initial information leads to rapid increases in quality, it might crowd-out the community's ability to develop ownership over the knowledge thereby reducing the propensity to make follow-on contributions and develop deep engagement in the long run.