CitizenLogistics

GROUNDCREW

www.groundcrew.us

A New Kind of Community Benefit Metric

It's an exciting time for community benefit metrics and related fields. New ways of thinking about value— including grant evaluation, corporate benefit, game systems, etc—are hot right now. These measurement techniques could reshape the global economy in profound ways, and it's important that we advance the state of the art. Measurement of community engagement and community benefit—which today involves laborious door-to-door before-and-after surveys like the Social Capital Community Benchmark or qualitative self-assessments like the B Corporation Survey, the London Benchmark Group, or Women's Funding Network's Making the Case—can be dramatically improved using location-based technology, mobile messaging, and social networks.

With these new technologies, we can measure the community impact of projects continuously, in real-time. In this document, we introduce a variety of *operational* and *real-time* community benefit metrics. We present metrics for (a) evaluating ongoing community health, for (b) measuring the real-time effect of community projects as they unfold, and (c) for evaluating the overall benefit reach for cross-community networks, foundations, and social initiatives.

Origins & Intention

Nonprofit social network *CouchSurfing.org* may have been the first to track these operational, real-time community metrics. Dashboards at CouchSurfing headquarters monitor values that keep the site effective and safe, while maintaining a trusting, sharing community with a strong offline experience (see sidebar, next page). The author was involved in developing these metrics for CouchSurfing.

Now, we have developed new metrics for another site, *Groundcrew,* focused on local and global communities. We aim to continue developing real-time metrics in collaboration with foundations, funders, and outside parties with an interest in evaluation and benefit tracking. This document is a starting point for that effort.

Rather than take theoretical positions about long-term "impact", these measures concern how communities and the interactions of neighborhoods and participants are changing on the ground *today.* For instance, whether the concerns that community members have are being attended to more often, or whether new opportunities are made available and taken advantage of. Theories of change based on distant results via, e.g., policy advocacy or some theorized cultural shift, cannot be evaluated using these metrics.

Three Kinds of Community Metrics

There is not one ideal measure of community health and improvement. Some metrics will be based on notions of public goods, such as connections between neighbors. These are easiest to measure for small projects. Others will be based on a concept of what it means to be a healthy community—for instance, whether a community can resolve issues that come up in a prompt and successful manner. These can be used to justify and explore the lasting impact and ongoing dynamics of projects. A third type of metric will be based on ideas about what makes a desirable global society, such as measures of overall happiness or thoughtfulness or opportunity. These can

Contact

Joe Edelman
CEO, *Citizen Logistics, Inc.*
joe@citizenlogistics.com
(413) 250-8007

Community Metrics, The Commons & The Next Economy

Our world economy is driven both by dollar metrics and by non-dollar metrics. In many cases, collections of assets—whether they are land trusts, foundation funds, car shares, human attention, or the resources of a business—flow in accord with non-dollar metrics. These include trust, reputation, and grant evaluation metrics, and they are used both as an alternative to allocating for pure financial return, and as a way to achieve lasting growth in a world where short-term financial metrics are rarely the best guide.

Increasingly, dollar return seems not to be a sufficient measure of what's worthwhile, and the call is out for a better, more accurate yardstick with which to run our communities and our world.

Software systems—especially social networks with real data about human connection and enjoyment —are accelerating both the monitoring of social benefit and the management of assets so as to maximize it. Indeed, algorithms like Google's PageRank and CouchSurfing's CouchSearch already allocate human attention and time in accord with notions of individual and group benefit. Sites like CityCarShare and NeighborGoods allocate physical assets in an automated, but non-financial way.

There are algorithms for measuring some kind of benefit and algorithms for allocating resources so as to maximize it. We want to explore this area more. This can shift our society from being driven by sales and dollar return towards a focus on long-term wellbeing for individuals, communities and the planet.

This shift from dollar-motivated enterprise toward commons-oriented approaches has started. It features pools of assets managed using a notion of public benefit. The vital thing is to choose metrics that don't imply particular tactics, but motivate broad, rigorous, socially-minded entrepreneurialism, decentralized decision making, and real-time response. We can build a system that's more efficient than traditional markets, that's focused on the common good, and that has rich rewards for everyone.

be used to guide long-term projects and to check assumptions about what the presence of public goods or community health might lead to.

An ideal system of metrics will relate, mathematically, these three kinds of metrics, and will justify each in terms of the others. This is what we will attempt to do.

Good metrics will also systematically avoid bias and externalities. They should take into account costs, and avoid privileging a certain kind of solution when another might be better or just as good. In particular we want to avoid the following situations:

Small Group & Demographic Advantage. We want to avoid the appearance of success when benefit is actually accruing to a small or a limited population.

Hidden Costs. When we count benefit but we don't count costs—including transaction, network participation, opportunity costs, and externalities—we can end up with misleading metrics.

Biases toward Shallow or Short-Term Solutions. We may need to explicitly model durational or long-term benefit in situations where present-time measures (like volume of sales per day or page views) lead to a bias towards short term thinking.

Metrics presented below will proceed from more naive approaches to those which take into account all of these factors.

The First Step is to Detect Experiences

Consider that any human or social metric is a quantification of personal or group experience. These numbers always relate to the experiences of individuals. With dollar metrics and eyeballs-on-websites metrics, we don't know much about what happened for the individuals concerned, besides that they were at the point of sale or at their internet device. But with notions of community health and global positivity, we want to know more about the experiences we are trying to generate, and we want to know if they are positive. In this section, we'll introduce several types of experience. In the next, we'll discuss relationships between them.

Experiences of Concern. When we realize an issue faced by someone or some group, we call that a concern. Concerns may arise in a local paper, a blog, in private conversation, or in private thoughts. Examples of concerns might be an elderly person who needs help, a school in disrepair, or the dreams or desires of neighbors which an individual wants to see realized.

Experiences of Availability. It is positive to realize that others are available to respond to a concern, or that there are resources available for an action. Examples include knowledge of cars to borrow, parks and meeting places to use, experts for consultation, etc.

Experiences of Action & Participation. Under the right circumstances, concerns and availabilities will become action and participation. A variety of technical and organizational factors act to limit or to enable this. Experiences of action are often social, can be positive experiences, and may lead to good will.

Experiences of Concern Resolution. When concerns are resolved effectively and with good will, we experience relief, triumph, satisfaction, etc. Even small concerns—like the desire to eat regularly—involve a positive experience when they are resolved.

Experiences of Good Will. When concerns are resolved with positive experience and widespread participation, experiences of good will seem to result. Good will seems necessary for sustainable community—for resources to continue in availability and for participation to be frequent.

Key

- Metric used by CouchSurfing
- Metric used by Groundcrew

What CouchSurfing Measures

- *Strangers introduced each hour by the system*
- *Which of those experiences are positive*
- *Which previously introduced strangers are becoming friends and/or continuing to communicate*
- *The ratio between online and offline engagement*
- *The levels of trust created in different areas and across different populations*

Piecemeal Measures of Community Health

Good Neighborhoods

- *Positive interactions and relationships with neighbors*
- *Neighborhood areas with less crime due to citizen activity*
- *Awareness of neighbors in need and neighbors taking action to meet them*

Engaged Citizenry

- *Diversity and participation in activity groups & events*
- *Neighbors reporting relationships with civil servants*
- *New, active dialogues with politicians & NGOs*
- *Citizens involved in maintaining public space*
- *Activities organized in public space*

Diversity

- *Relationships formed between neighbors from disconnected ethnic and social groups*
- *Diversity index of populations using public space*

Sustainability

- *Participation in local solutions for food, recreation, child care, etc.*
- *Reclaimed public space due to community care.*

Thanks to:

C Fenton, J Stone, K Souren, *CouchSurfing*
M Patel, *Knight Foundation*
T Mollner, *Calvert Investments*
J Porcelli, *Neighbors for Neighbors*
J McGonigal, *Institute for the Future*
J Greenhall, *Santa Fe Institute*
C Thornton, *Citizen Logistics*

Records of Experience in a Society

At any moment, a certain percentage of Earth's population are having at least one of the experiences above. These experiences are detectable, countable events, and can be analyzed in terms of their value, their cost, their frequency, and the equity of their distribution. Furthermore, these experiences connect person to person across social and information networks—one person's availability relates to another person's concern, good will is the consequence of participation, etc—and they also relate to assets available and to the provision of goods and services which sometimes have a dollar equivalent value.

Keeping more-or-less comprehensive records of experience is only recently possible. Between web server logs, real-time analytics, thumbs-up or 5-star ratings systems, mobile location data and lightweight follow-up mobile survey questions, it is possible for any community project that has a significant online/offline component to keep records of experience. These records form the raw material for a new family of benefit metrics.

Theoretical Models

Some of the metrics below are relatively theory-free, relying mainly on reports of positive experience and the assumption that these are good for people. The more sophisticated approaches, though, rely on models and assumptions. By making simple models of these aspects of community, we can start to diagnose and account for problems on a finer scale.

Models used here are based on two related notions: *optimal empowerment* and *healthy resilience*. The former concerns the levels of attention and participation in a community, relative to welfare of members. The latter concerns how well a community resolves group and individual issues that arise. The intuition here is that a healthy community attends to the potential of its members and to community issues that arise—that these, taken together, form the true interests of the group.

To formalize our notion of healthy resilience, we claim that a community is healthy if, as *concerns* arise within the community, they are *responded to* and *resolved* in a good way. More specifically, that community ideas are turned into action and participation, that community resources and expertise are leveraged and made available, that concerns and ideas are published, that participation in the resulting endeavors is achieved, and that resolutions are found which stakeholders feel good about.

We can represent this with the pipeline of conversion rates that forms the basis for the "Diagnostic and Operational Metrics" section below:

Figure 1. CAARG Pipeline

CONCERNS → AVAILABILITY → ACTION → RESOLUTION → GOOD WILL

Later in the paper, some metrics rely on an additional notion of of *optimal empowerment*: a society is optimally empowered if, there is no subgroup which lacks human participatory resources (attention, coordination, and participation, whether pertaining to their needs, their dreams, or their potentialities) which would increase their long-term positive experience or their availability to others.[1]

Metrics

We will now present several approaches to calculating social benefit. In the first section, we do so by counting positive experiences. In the second, we assume a certain pipeline that generates positive experiences, and evaluate the health of that pipeline. Later sections make adjustments for bias, for opportunity costs, and to account for different measures of value.

Counting Experiences

Concurrent Positive Experience  

For any moment of the day, some people have reported experiences of good will, or that concerns were being resolved in a way they feel good about, or that they've participated in an action and liked it, or that they've been set at ease by the availability of community resources. We can simply gather up all this positive experience and talk of the concurrent positive experience in a community or around the globe. We track this at CouchSurfing, and, as you might imagine, it's a lovely chart to watch.

HOW TO CALCULATE: With both CouchSurfing and Groundcrew, we ask participants to rate the quality of their experience, and collect information about their duration. We also collect information about how much non-positive effort on the site is necessary before our average user has some positive in-person experience with another user. Finally, we know where our users live and often know how many non-CouchSurfing friends they have. This data lets us calculate transaction and opportunity costs and get a bigger picture of the marginal utility of CouchSurfing in terms of how much net new positive experience we bring to the life of an average user, and of our users in total.

What are the problems with this approach? It can hide problems with unproductive or counterproductive positive experience, with biases and inequities, and with positive experiences that displace even more desirable, experiences. The more elaborate measures, towards the end of this document, take all these costs and biases into account.

Dollar Replacement Value for Services Enabled  

Another way to look at concern resolution and participation is that they correspond to services rendered and to rentals or labor costs, respectively. Considering each act of human assistance as a service rendered, and each durable good employed as a rental—what would be the dollar cost to replace these community assets at market rates, from nearby vendors instead of through community and kindness?

This is a compelling measure, as it allows for a direct representation of Social ROI in dollars. CouchSurfing calculates its dollar replacement value for lodging and matchmaking at around $120M/year, which is delivered for an actual yearly cost of $1M. It provides a social ROI of 120x on dollar investments, year after year.

There are two drawbacks with using this as a primary metric. First, it doesn't take into account any part of the pipeline beyond participation. For example, the generation of good will is not accounted for, nor whether the participation or service resulted in a resolution of concerns. Second, this metric has a bias towards high-dollar value services instead of towards high positive experience services.

HOW TO CALCULATE: The various experiences and activities enabled by the social network need somehow to be correlated with the kinds of labor and service that take place on the market. This is a lossy conversion, since what community members do for their neighbors is never identical to what's available in the market. It also requires a geographic understanding of market rates. With CouchSurfing, we use hostel rates, entertainment services, and matchmaking services.

Concurrent Positive Effectiveness
A straightforward next step is to remove positive experiences that leave the pipeline before reaching concern resolution and community good will. Someone may enjoy reading about a concern but do nothing about it. A group may come together in enjoyable participatory action that leaves a neighborhood just as opportunity-poor as before. To get a better picture of how heathy a community is, we can keep track of our conversion rates along the pipeline, and multiply our recorded positive experience rates by the likelihood they will convert down the line, before adding them together.

HOW TO CALCULATE: This approach requires better data than CouchSurfing keeps about needs/concerns. We need longitudinal data about different concerns that were expressed and whether they reached each pipeline stage in turn and/or were eventually resolved by what happened. With Groundcrew, or with other citizen issue tracking systems like SeeClickFix, this additional data is easy to collect and monitor.

Concurrent Individual Opportunity
Another approach focuses on the longer-term, lasting experience of ambient availability. Through CouchSurfing, a certain number of users know that they can travel at a dramatically lower expense, and that they will have ambassadors and new friends to connect them with social scenes and ways of living in every tourist city on Earth. Some of these people would have felt stuck or disconnected without knowing of this availability. For many it was a revelation when they found out, and resulted in a change of lifestyle and mindset.

It's a worthwhile question to ask: how much of this experience of opportunity and availability are we able to provide? This is not just a question of where we have hosts, but also of where people are searching, where they want to travel, and how many of our hosts are burned out... our excess capacity, its distribution relative to the places people care about, and its accessibility to the people that care.

HOW TO CALCULATE: A straightforward approach is to keep track of what happens when people search, how often they search, and how often, when they follow up with search results, the apparent availability turns out to be real. With CouchSearch, we continuously monitor how easy it is to reach an available host in different areas, the expected quality of interaction with those hosts, and the expected level at which needs will be met by those hosts. We multiply these together and take a sum across active users.

Diagnostic and Operational Metrics

Measures of aggregate experience are useful, but you need to get more specific in order to diagnose particular community failures, or to operationalize more targeted projects. Here we present measures focused on particular linkages in the pipeline of healthy community concern resolution. Instead of experience counts, here we measure conversion rates and time deltas from one kind of experience to another (possibly broken down by demographic). In one special case— when describing the size of the community assets that are being brought to bear—we measure in dollars per capita.

Time to Positive Resolution (and related)
How long, on average, does it take for a concern of a certain size to be resolved with good will? How does this vary across concern sizes and across sub-demographics within the community? What percentage of concerns are resolved? And how many concerns can be resolved simultaneously?

HOW TO CALCULATE: There is a strong parallel here to efficiency measures for telecommunications networks: bandwidth, throughput, goodput, latency, packet loss, etc. Generally, the same analyses can be used. Demographics differences can be represented as subgraphs, or, for or a more sophisticated view of interrelations between demographics, a combination of social networks analysis and network flow analysis can be used.

Publishing Efficiency
Moving concerns along this pipeline and generating good will requires communication and connection at each stage. Questions like the following arise: (a) Do those who are available know about the concerns? (b) Do those who are acting know enough to act effectively? (c) Does the community know enough about what happened to generate good will?

HOW TO CALCULATE: Publishing can be considered as a process of allocating community attention—conscious of opportunity costs outside the model—across a variety of topics so as to maximize positive effectiveness (in person/time or with a dollar value). It is important to account for the cost of attention that doesn't generate action, resolution, or good will. If newspapers and websites were to make this calculation, some community-themed websites and papers would discover they spend more time in distracting community members from being of service than in motivating it.

Dollar Replacement Value for Asset Availability
Without shared assets and shared expertise, it's hard for a community to resolve concerns. Indeed, the mere availability of assets and experts is a source of positive experience (see Individual Opportunity, above). These assets and this availability can be given a dollar value, much like Dollar Replacement Value for Services Enabled, above. This values a community like a company with cash, assets, and human resources on hand.

These numbers can be quite high; in the case of CouchSurfing, we are talking about year-round availability of more than 1M guest rooms in cities everywhere. Replacement cost to maintain this availability would be much greater than the early estimate of $120M to replace only the actual use.

HOW TO CALCULATE: This metric considers what was available, rather than what was used. Across all listings, what is actually available to the majority of users? Across those shared goods, what is the total value of these assets? Across all services provided (including expertise, mutual aid, etc) what would their value be when considered as independent businesses?

Hidden Costs

Metrics based solely on counting or valuing positive experiences may be ignoring hidden costs. These can be grouped as transaction costs and opportunity costs.

To say that a community project has caused a positive experience — the resolution of a concern, or the creation of good will — would be inaccurate without accounting for the fact that any action or involvement in the community would replace a different action or involvement if the project wasn't there. This is the opportunity cost. With CouchSurfing, each member has relatives and friends they travel with and enjoy outside of our system. To the extent that our system takes time away from this pre-existing conviviality or simply substitutes it with a positive experience of a different kind, we cannot claim we've created net positive experience.

Similarly, there are time costs that aren't positive or community building associated with the use of any system, including costs of signup, of reading about others' concerns, reading through listings, search, etc. These are transaction costs.

When we account for these costs, we register *net* positive experiences—those experiences provided minus those displaced.

Net Orchestrated Positive Experience ©
To measure positive experience as a net value rather than a gross, one needs to know both the transaction and the opportunity costs. For the former, we measure the time and attention participants provide where they are not engaged directly in a positive experience. This includes negative experiences as well as time spent signing up, browsing, getting trained, etc. For opportunity costs, it's necessary to have an understanding of what, for a typical participant, engagement with the service is a substitute for, and how positive that other engagement would be. The marginal utility or net positive experience of the service is the additional positive experience across all participants.

HOW TO CALCULATE: With web services and/or GPS data, we have the advantage of being able to track almost all time involvement for users. A service like CouchSurfing or Groundcrew has a value proposition, and transaction costs include any time spent that's not related to that value proposition. Both transaction cost time and service-related positive experience displace a baseline of positive experience present without the system which can be explored through signup surveys and correlated with social network sizes and demographics.

Measures of Burnout ⓖ
Burnout is another cost of operating a system which sends requests or puts information in front of people. Burnout happens when transaction costs exceed utility: unless community members receive just the right number of requests —especially requests that are meaningful to them, that they can act on, and that engage them—burnout will result.

HOW TO CALCULATE: Three common types of burnout are: when a small cohort receive too many requests; when they receive requests which are not engaging enough; and/or when they do not receive sufficient or accessible enough requests to engage them in the first place. The first two conditions can be measured by segmenting a user population by signup date or stage, and measuring decreases of conversion and engagement over time. The latter can be detected by looking for correlation between cohorts that have been less involved and that are not practically available. To the extent this is happening and concerns exist, failure to engage this community can represent a deadweight loss.

Long-term Positive Experience

Another issue with using real-time direct counts of positive experience and time-to-resolution measures is that they can privilege short-term solutions over long-term solutions that might take more time or attention up front but would lead to more positive experience or availability in the long run. In this case, it can make sense to discount short-term-biased experience from our measures.

Adequately Thoughtful Effectiveness ⓖ
We can define a concern resolution as *adequately thoughtful* if no reasonable amount of additional consideration or planning would have lead to a significant increase in positive experience for those affected over time. This measure then starts with Concurrent Positive Effectiveness but then discounts all positive experience that was nominally effective but not adequately thoughtful.

HOW TO CALCULATE: Calculating this metric requires longitudinal data about individuals and groups and a very large, active user base. The trick is to continually run experiments across a range of very different kinds of concerns where a control group gives an unmodified amount of consideration to the concern, and an experiment group is lead to give additional consideration. We then determine a correlation between additional thoughtfulness and long-term differences in positive experience[2] and can discount our Concurrent Positive Effectiveness suitably.

Discounting Demographic Advantage

Funders, evaluators, and project managers will want to know if a community project is mostly beneficial to a small part of the community or to a certain demographic. In some cases it may be desirable to discount positive experiences or asset availabilities where advantages are localized. There are several ways to do this.

HOW TO CALCULATE: There are three different approaches, each with its own strengths. The simplest, if you have demographic concerns about race, gender, age, and the like, is to measure differences in experience or asset availability across these bins. This is straightforward, but assumes knowledge of the breakdown. An alternate approach is to use a Gini Coefficient to measure overall inequity in what your use population has received. The third, and most interesting approach, is to measure not just equity, but *commingledness*—whether different groups are interacting or whether there are many isolated demographic or use-level cliques. We can do this with a social network analysis and a measure of network flow into and out of a wide sample set of nodes.[3]

It's important to note that these measures are mainly necessary if you are counting the people who are served by a project. An alternative approach, explored in the next section, is to count the community members who *weren't* served. More specifically, you can measure the distance between your community's observed functioning and some notion of ideal function in which everyone is served.

The World We Want to Create

Our hopes for humanity do not end with positive experience and concern resolution, although these are important and very accessible measures. Beyond these are notions of empowerment, stewardship, and care. There is the ideal of giving every person and every possible group the awareness of their true interests, dreams, and desires and of providing the full extent of support to make them a reality.

As with notions of optimality and efficiency in economics, these notions are useful even if we can never achieve this optimal state. Notions of distance from the optimal state are useful. The raw materials for calculating distance against this ideal are, again, records of positive experience and concern resolution.

Distance from Optimal Empowerment ⓖ
Proceeding from the reasoning above about ADEQUATELY THOUGHTFUL EFFECTIVENESS, we can ask a simple question: given observed correlations between thoughtfulness and coordinated participation and long-term benefit differences observed, how much additional thoughtfulness and coordination would be necessary to consider and coordinate around every individual and group to the point of negligible or diminishing returns?

Limited-Budget Distance from Optimal Empowerment ⓖ
Alternatively, you can base this calculation on how optimal your distribution of attention and participation is, assuming a limited budget of human attention and participation which can be spent in a way that is effective and generates positive experience and good will. In this case, it's the amount of thoughtfulness and coordination that is misallocated or under-deployed which is measured.

HOW TO CALCULATE: The limited budget case is complex computationally even with good data about the relative effects of attention in different places on positive experience. But approximating a rough value using heuristics and binning is not hard. In the unlimited-budget case, the value is a straight sum of projections of additional attention that would be effective across individuals and some space of reasonable subgroups to account for.

Further Work

In a future document we will recommend particular metrics from this collection for particular kinds of projects. Some will be appropriate for small projects, others for large networks of funders, investors, or government initiatives. Teasing the various factors apart will be a separate task.

We also have metrics developed for a variety of special cases, including the maintenance of safety in social networks, the building of trust and good will in communities, the conversion from online to offline relationships, the running of community app contests and social sharing systems, and a variety of municipality-themed metrics that related to those discussed. These too will have to wait for future papers.

Finally, it remains to relate these metrics numerically and with observation to the leading measures in the field, such as the Social Capital Community Benchmark, and to document the deployment of these measures in a set of case study projects and communities.

Conclusion

In our society, the easiest way to bring resources to a project is to sell something. Sales are straightforward to measure and carry an immediate resource allocation. Unfortunately, generating sales and providing long-term benefit are not the same thing. These two diverge most with regard to services that elude commodification (love, attention, help), that are abundant (mesh wireless access, water, mp3s), or that are difficult to gate-keep (public parks, the planet, other commons). These services end up being under-valued, under-incentivized, and made artificially scarce.

It's important to think critically about any bottom line metric, whether social or financial, and it's dangerous to embrace any formula too tightly. The promise here, though, is that these new formulas might help us disentangle ourselves from a previous embrace: our embrace of financial and sales-oriented metrics.

The further we can push our measures along the time axis from the point of *sale* to the point of *long-term benefit*, the closer we can come to allocating resources for the greatest benefit.

---

[1] This is similar to the notion of Pareto optimality, but emphasizes invention and care instead of trading or reallocing excludable goods.

[2] It is difficult to say what's the best way to detect a difference in long-term positive experience records between groups. Rather than taking an average or a median rate of positive experience, I would suggest using a multi-component analysis of time-scale data, such as a set of one-dimensional moments, and then recognizing any significant change in values that could be called an increase as evidence that additional thoughtfulness had an effect.

[3] Similarly to the network flow measures used in attack-resistant trust metrics like Advogato.