

Towards a better measurement of social influence.

The Premise:

The measurement of social influence is fundamentally flawed due to its reliance on a limited data set. An alternative is therefore required.

Gauging the true influence of an individual is a complex affair and hinges on a host of factors including a stable identity. A consistent identity across services brings with it an inherent element of trust – people know who they are dealing with as opposed to someone using a pseudonym. I will address identity as a part of influence later.

Current systems such as Klout and Kred do not measure influence, they measure interaction by way of pure numbers: shares, retweets, likes, +1s, comments etc. extrapolated out over your connections and those sharing your content. This is not true influence and, in treating it as such, we can become slaves to the gamification of activity.

Where to measure

I believe that influence cannot be accurately measured externally from the data source as there is a limit to what can be gleaned from what is publicly available. The service hosting the data (be it Twitter, Facebook or Google) has a better understanding of exactly what happens to that data including other factors such as link tracking.

Activity can combine positive reaction, click-throughs and even actions at the source material and we should also consider the quality or sentiment of responses. Measurement based on activity alone, however, presumes we must maintain a constant level of activity to stay influential or even increase it in order to exert greater influence.

We can move on to examine two further factors which affected influence in timescale and context.

Klout, for example, uses a 90 day period to calculate influence but a couple of days of inactivity will cause busier days to drop out of the "influence range" with an almost immediate impact on influence score. While it is reasonable to expect someone who is more active to have more influence it doesn't follow that this is always the case.

Using the analogy of a scientist producing research papers, as an expert in their field they will be highly respected and have influence but they will not be releasing papers on a frequent basis. As long as the research is still relevant and has not been superseded their influence will not diminish.

Influence will, in many cases, decrease with time but a reputation that has been built over months or years will not immediately start to erode after a degree of inactivity.

Reputation is, therefore, just as important as activity, perhaps more so in certain scenarios. Being known as the go to guy or expert in the field speaks volumes and should have more impact than the number of "interactions" on any given day. Klout tries to address the area of reputation by giving users the ability to award +K's for particular topics. Kred will give extra points for additional activities including those outside of social networks or even offline but these don't appear to be primary drivers; they may give us a slight boost but we are still largely dependent on activity.

Creation v Curation: context

Not everyone uses social in the same way but judging influence by activity treats us all the same despite our circumstances differing considerably.

Is a creator or through leader more influential for supplying the topic of conversation or is the curator more influential for spreading its reach? Obviously, the two go hand in hand but the curator could not act without content.

One argument for accepting the current systems is that they are the same for everyone and are, therefore, fair but do we *really* get a meaningful comparison? Is the same method of influence measurement suitable for both the creator and curator? If not, how do we distinguish between the two?

The proposal:

As can be seen, the current methods of influence measurement have inherent difficulties which leave us with an inaccurate picture. So, how do we fix it?

The 3 R's of social influence

As well as current activity based measurement there needs to be a more qualitative approach to influence which, I believe, will have an equal - if not greater - impact as the quantitative measurement. Combined, this will create the 3 Rs of influence model:

- Reach
- Reputation
- Relevance

Reach

Irrespective of anything else, our prospective audience is still important for calculating influence - the more people we have an impact on the more influential we are. This is what is measured by the current systems.

Our reach is a combination of our direct first level connections and those secondary connections exposed to our content by re-shares, retweets, etc. Klout breaks this down into three components:

- true reach - the number of people influenced,
- amplification - how many people respond or share your content, and
- network – the influence of those in your true reach (the ability of your connections to influence the secondaries)

The number of connections alone is not necessarily a factor when gauging influence but we must consider changes in this value when looking at influence. I believe "trigger" events such as gaining or losing a number of followers over a certain threshold could also prompt for a re-evaluation of influence.

If someone has suddenly lost hundreds of followers could this be because of a change in their credibility? Perhaps they have said or done something detrimental or inappropriate. Conversely, gaining hundreds in a short space of time will probably indicate a recommendation by an influencer - such as being included in a shared circle or mentioned by a big hitter on Twitter - or the possibility that the person has published influential content.

Follower numbers are fluid so gaining and losing is an expected part of partaking in social networks but finding the reason for a sudden change and discovering the sentiment surrounding that event can shed extra light on influence at a given point.

The number of people we influence is obviously a major factor but numbers alone can give a misleading impression, we also need to know *how* we influence them to gain a true picture.

Reputation

As has been alluded to above, a person's credibility or reputation will be a factor in determining the degree of influence. Who are they? What do they do? Are they a subject matter expert and is their opinion trusted? It will itself be affected by identity and sentiment.

A person's trustworthiness can be viewed in general terms but their actual influence will differ depending on the topic in question. While Klout may decide which areas we are

considered influencers for we still have a single score that might give an incorrect impression.

How can we determine reputation?

As I have mentioned, influence cannot be accurately measured externally from the data source as non-public data will not be factored in to the equation. The data host is best placed to properly analyse what they hold and Google is possibly best placed to gauge influence due to the combination of social signals and search data at their disposal.

Identity

A strong reputation is dependent upon a consistent identity and the web has been crying out for a standardised identity system for years and I believe that it has to come down to one of the big hitters to be able to pull it off. The open source efforts such as OpenID are never going to catch on with the general populous as there is often a lack of trust in the open source movement.

With Google's privacy efforts etc. under observation for the next 20 years and recent advances with Plus and social search all in house there's probably never been a better opportunity to standardise online identity. Google+ is intended as such an identity service.

AuthorRank would appear to be a good starting point for calculating reputation; Google's "Agent Rank" patent outlines a method to use "*a reputation score associated with the maker in calculating a quality score for the document*".

It is intended for AuthorRank to influence the traditional PageRank to improve the accuracy of our search results but to do this the system will first need to rank the author. According to the patent "*an agent's reputation can be derived using a relative ranking algorithm*" with an author being given a higher reputational score if their content "is frequently referenced by other agents or content" providing it is signed by the author. The Authorship scheme which connects to Google+ using *rel=author* provides the method of identification and, thus, the signature that can be applied to content.

By recognising the *person* responsible for content and their relative reputation based on external references we can assess the influence of that person based on the perceived value of the content. The external references will carry their own weighting as some may also be attributed to an author with their own reputational ranking; a source with a higher reputation will, therefore, have a greater impact.

Context

AuthorRank has an additional advantage in that it helps us distinguish between the different types of influence enjoyed by creators and curators.

A curator can have reach and relevance and also some reputation via sentiment (+1s etc.) for consistently sharing interesting content but will not (and should not) have AuthorRank.

Allowing us to make this distinction is a vital step in assessing relative influence and gauging if a person is a contributor or a traffic driver. We can treat different types of people in different ways and, perhaps, curators will be encouraged to provide original thought and actively discuss the links they provide.

Sentiment

As mentioned above, we should be considering the quality or sentiment of responses to any content - as well as their number and source – to gauge their impact on influence. At present we have simple methods of collecting positive sentiment via the count of likes and +1s but this is not sufficient.

A lot of work is being undertaken in the arena of sentiment analysis as it applies to social networks with a focus on the emotional response to content and events. This is obviously of great benefit in calculating the reaction to content as opposed to merely recognising it has been referenced.

By classifying the polarity of external references at the document or even sentence level we are able to examine the context in which our content is quoted and the subsequent impact this will have on AuthorRank. Being linked to, or quoted by, a recognised influencer will potentially give a false boost to AuthorRank unless we know whether the reason for that link is positive, negative or neutral. Being discredited by a proven expert should serve to decrease reputation and AuthorRank.

While there have been calls for standards in sentiment analysis this is perhaps not essential as long as there is consistent measurement; in the instance of a single agency (such as Google) performing sentiment analysis for influence this is obviously going to be the case.

In this context both AuthorRank and sentiment are helping to establish a trust for the individual and their content or actions.

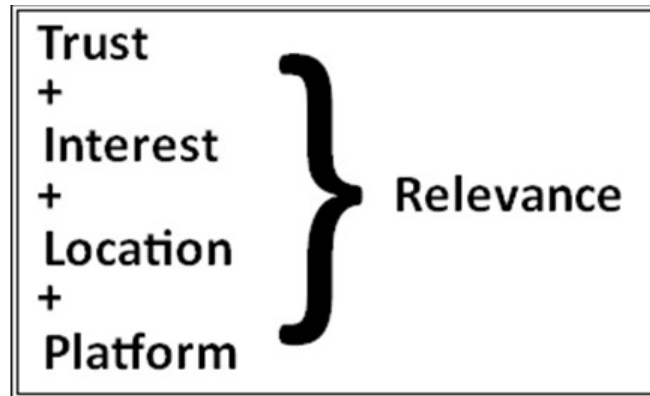
Relevance

We are more likely to be influenced by those who are relevant to our circumstances. Relevance is not static and will depend on our interests, location and even current platform – be it web, desktop application or mobile.

Social signals on their own have the same problem as merely counting links, there needs to be an identifiable element of trust which is, I believe, one of the biggest reasons that Google initially launched Google+ with a real names policy.

With the launch of a new social product, that also aimed to be an identity service, Google needed to establish a “trust network” as well as a social network in order to be taken seriously.

One big movement in social which has yet to come to fruition is the interest graph. The big players are all dabbling on the periphery but so much more can be done to truly utilise our interests to create connections and ad hoc social graphs as opposed to the explicit graphs we create by following “people”.



The start-up [mon.ki](#) aims to bring some social relevance to your browsing, what is described as your “social compass”, by automatically pulling out the profiles of those associated with the content you are viewing as well as an amount of social reaction. It is immediately obvious that there is massive potential with such a concept but it could be extended further.

Social relevance can easily act as an indicator of influence - by keeping track of those associated with the content we view patterns can be established from which we could possibly infer influence and, indeed, relevance.

Combining AuthorRank and social relevance we can attack the same issue from both sides:

- AuthorRank will establish influence based on how Google sees the author as an expert based on topic and peer recommendation, while
- social relevance will determine degree of influence due to the connection between author and reader especially if across multiple sites and sources

Tim Delhaes, mon.ki CEO, has confirmed that “degree of influence” is an area the company expects to target.

An outfit such as this would be an ideal acquisition target for someone like Google or Facebook by allowing them to extend their knowledge of our connections beyond the confines of the social based on our consumption habits.

Repetition?

While it has been suggested that a fourth R could be Repetition I feel that this would give the wrong impression of what is required.

Consistency of quality is obviously required to remain influential but not required in all cases (see the example of our scientist). It is, therefore, an issue of semantics in that repetition implies a need to constantly post and thus falls in to the same trap as solely relying on activity figures for our measurements.

Consistency of quality is inherently implied by trust and reputation and, for repeat creators, that quality is more important than the actual repetition of releasing content.

Conclusion:

Social influence is complex and to get even a moderately accurate picture we must gather input from numerous sources.

Influence can be seen as the sum of the 3 R's: reach, relevance and reputation, which in turn can be gauged using a combination of trust, interest, location, platform, sentiment and AuthorRank.

Time and again we keep coming back to one vital element in trust so it can be seen that creating a reliable "trust network" is key to establishing influence.

The diverse nature of these signals, and the need to collect them in different locations and in differing ways, indicates that a third party remote from the data is unlikely to be able to achieve accurate measurement. It would, therefore, fall to a major player such as Google or Facebook with a wide reach and unrestricted access to varying sources of data.