What is in PageRank? A Historical and Conceptual Investigation of a Recursive Status Index.

Abstract: This paper proposes an analysis, based in a software studies mindset, of Google's *PageRank* algorithm. It develops two lines of investigation: first, it situates this 'evaluative metric' in a larger genealogy of ideas, concepts, theories, and methods that developed, from the 1930s onwards, around the fields of sociometry, citation analysis, social exchange theory, and hypertext navigation. This backdrop is presented as a conceptual *a priori*, in the Foucauldian sense, from where *PageRank* becomes 'sayable'. Second, by comparing the algorithm to a close cousin, Jon Kleinberg's HITS, and by examining a particular parameter in the *PageRank* model, the paper shows that the concrete model does not follow teleologically from this historical *a priori*. Behind the particular empirical case lurks the larger epistemological and methodological question of how to study software. The paper advocates a multilayered approach that combines different types of methodological and conceptual resources.

1. Introduction

When asking the question 'what is an algorithm?', computer science offers numerous definitions, such as Knuth's classic 'a finite set of rules which gives a sequence of operations for solving a specific type of problem'[1]. From there, we are frequently directed to the underlying principles of mechanical computation, and to the foundational work of Turing, von Neumann, Church, and other mathematicians. But what, we may ask, is the 'problem' referred to in this definition? At another point in Knuth's pedagogical introduction, we learn that algorithms involve 'some *action* to be performed or some decision to be made'[2], but, again, what does the 'some' stand for? Because, ultimately, if our business is not with the matter of how mechanical computation is possible in the first place, but with *software* as an object in-the-world, the question to ask becomes 'what is in an algorithm?', that is, what is the 'problem' and what does the 'some' refer to? The elegant concept of computation then quickly begins to *bloat up* with many different things: real computers, not just abstract Turing machines; real software, lodged in tight networks of *other* software, all written for a *purpose*; knowledge, ideas, skills, tools, methodology, habits, and values that permeate *practices* embedded in layers of social organization, cultural configurations, economic rationales, and political struggles. If software, in the words of Netscape co-founder and venture capitalist Marc Andreessen, is 'eating the world'[3], then the world – culture, economics, politics, sociability, education, desire, and so forth – *fills up* software in turn. Understanding how software is in-the-world and how the world both pushes into software and is captured by it, is one way of articulating what the emerging field of *software studies*[4] is trying to achieve, and this paper attempts to contribute to that project.

It does, however, not intend to put forward a consistent theory or a polished methodology; instead, it proposes a somewhat *naïve* case study of a specific algorithm, Lawrence Page's *PageRank*. Despite this narrow focus, I hope to develop a viable approach to studying software as well as to illustrate the larger category of what could be called 'evaluative metrics', calculative procedures that aim at assessing importance, quality, relevance, performance, and so on – and in particular those relying on techniques located in graph theory and matrix algebra. These metrics have come to play a significant role in a growing number of settings, especially in the large-scale information systems that proliferate on the Web, where they are used to filter, select, order, attribute, distinguish, and connect. The Google search engine is indeed the tip of a much larger iceberg. The choice of this specific case is further motivated by the immense influence the network model has gained in many scientific disciplines, which can be explained, at least partially, by the considerable means for calculation it provides. Finally, *RageRank* is one of the few algorithms venturing into deeply mathematical territory that has not only been studied extensively by mathematicians and computer scientists[5], but is also regularly commented on by scholars in the humanities and social sciences[6]. This relative familiarity with the concept and its application in Web search will hopefully make certain aspects of my argument more vivid and provide grounding material for the more abstract technicalities I will address.

While this article can also be read as a contribution to the field of *search engine studies*[7], I am limiting my analysis to the *PageRank* model as it is laid out in two papers[8] and, more importantly, two US patents[9] that are, interestingly, more explicit in their citation practice than the academic publications. I will make only passing reference to the Google search engine. This focus allows me, on the one side, to concentrate on a model that is both well documented and well studied, and, on the other, to treat it as an example for a set of techniques that are used far beyond the domain of Web search. Furthermore, instead of discussing social or political *effects*, I want to examine the logic it encapsulates and the conceptual *commitments* it embodies. My main goal is to show how a multilayered yet contained reading of a very specific computational artifact can produce a nuanced account that is attentive to 'cultural logics', but does not *dissolve* concrete technical concepts and decisions in a homogeneous and homogenizing logic of 'computationalism'[10]. While I do believe that the network approach to evaluative metrics implies a reductionist and often totalizing conceptual horizon, I will argue that concrete algorithmic objects do not follow teleologically from this horizon, that there are 'margins', and that these margins are far from insignificant. In this context, a historical investigation provides valuable help in coming up with a richer account than a mere technical reading could provide on its own.

The intellectual genealogy of the models and methods in question here can be traced back to at least the 1930s and leads to an important *trading zone*[11] between mathematics and the social sciences that is largely unrelated to the well-documented history of a similar space of conceptual exchange, that of statistics[12]. In the first part of this paper, I reconstruct certain aspects of this historical context and focus on the particular idea of measuring the 'status' of a node in a network *recursively*, that is, taking into account not only the number of connected nodes but also their respective status.[13] This idea is generally presented as a core innovation in the *PageRank* ranking model: when calculating the 'importance' of a Web page, a link from a highly ranked site is 'worth' more than a link from a site with a low rank. But if we do not believe in invention as creation *ex nihilo*, we may want to investigate the 'place of emergence'[14] from where the model became *sayable* and, consequently, 'calculable' and 'programmable'. This refers, of course, to what Foucault calls the 'archive', a concept aimed at defining a 'historical *a priori*' that does not simply result from a history of reason or of the economy or of the state, but is made up from the complex relationships between heterogeneous elements, between words and things. These relationships are neither linear, nor amorphous or random but riddled with intricate regularities and distinct figures, forming systems that define *the law of what can be said* – 'la loi de ce qui peut être dit'[15] – in a given discursive practice. The concrete forms these systems of rules take are not merely the imprints of transcendental historical forces, but the result of complicated and contingent exchanges between contained internal systematics and the histories of other discursive practices. While the reconstruction of the full *archive* of evaluative metrics based in graph theory is well beyond the scope of this article, even an archeological *probe* allows us to situate both interpretation and critique in a space where internal and external trajectories can be kept in balance.

The second part of the paper examines *PageRank* more concretely, first by comparing it to the very similar HITS algorithm[16], and second by closely scrutinizing the 'damping' factor α with the help of the sociometric texts that first introduced such a parameter into recursive status calculations. The methodological question here is how far we can analyze algorithms as cultural objects *without*, and this is crucial to my approach, dissolving models into metaphors. While it is probably impossible to uphold a strict difference between the two terms beyond a certain point, the intellectual domains in focus here – mathematics and computer science – are built around the notions of logic and calculability and this constraint implies certain *forms of reasoning*, and in particular the expression of ideas as formal models. Instead of 'liberating' the *PageRank* concept from these constraints, I am interested in understanding how the network description frames and formalizes social phenomena in a way that *complies* with them. Between a broad idea for an evaluative metric, its expression in mathematical notation[17] as

$$r(A) = \frac{\alpha}{N} + (1 - \alpha)\left(\frac{r(B_1)}{|B_1|} + \ldots + \frac{r(B_n)}{|B_n|}\right)$$

and the computational method written in a programming language[18], a number of transformations have to be performed to arrive at a form that satisfies these constraints. But the *shadow of computation* – the often implicit requirement to only express ideas that *can* be made computable – already constricts the space where to look in the first place. *Modeling* is not simply the process of creating a formal version of a more general idea; it is

in itself an idea of how to meet the world.

Treating *PageRank* as a model also means that we can *run* it, explore what it does, feed it with data, tinker with it, and so on. At the end of this paper, I will quickly show how such a 'hands-on' analysis can look like and what we can hope to learn from it. The first step of my argument, however, consists of a closer look at the history of graph theory and its close relationship with the social sciences.

2. Social Theory, Graph Theory

In the introduction to his *Theorie der endlichen und unendlichen Graphen* from 1936, the first textbook on graph theory, Dénes König wrote that '[p]erhaps even more than to the contact between mankind and nature, graph theory owes to the contact of human beings between each other'[19] and thus points to the often overlooked fact that modern graph theory developed, perhaps even more so than statistics, in close contact with the social sciences. Rather than being imported into these disciplines as a ready-made mathematical toolkit, graph theory was shaped in no small part by *the measuring fever* in the field of 'small-group' research, founded by Kurt Lewin and Jacob L. Moreno. Even the mathematician Frank Harary, the central figure in the development, application, and standardization of graph theory in the second half of the 20th century, first came into contact with his future field of specialization[20] at the *Research Center for Group Dynamics*, founded by Lewin in 1948 at the University of Michigan. This close relationship warrants some further investigation.

2.1. The Development of Graph Theory in Relation to the Social Sciences

Despite its enormous success in recent decades, graph theory – 'a mathematical model for any system involving a binary relation'[21] – has historically not been considered a 'noble' endeavor. Although prominent names such as Leonhard Euler, Gustav Kirchhoff, Arthur Cayley, and William Hamilton are attached to its history, it is only in the second half of the 20th century that it emerges from what J. H. C. Whitehead called the 'slums of topology'[22]. That the *Journal of Graph Theory* was only founded in 1977 further testifies to a lack of interest from (pure) mathematicians. But besides the growing number of practical applications in chemistry, physics, and engineering, it is in the social sciences, and in particular in social psychology, that a *problem space* appears that proves particularly fertile for the growth of a mathematics that deals with structure.

A first line of development can be traced to Jacob L. Moreno's *Who Shall Survive?* from 1935, an original and idiosyncratic volume that launched the field of *sociometry* as 'the mathematical study of psychological properties of populations'[23]. Developed out of group psychotherapy, sociometry was essentially organized around the *sociometric test*, a questionnaire distributed to small- and mid-sized groups – schools, factories, etc. – that asked people to *choose* the individuals they liked best, admired most, had the most contact with, or a similar question. The resulting (network) data was thought to reveal the 'psychological structure of society'[24] and was displayed and analyzed as a *sociogram*, a manually arranged 'point and line' diagram that would be called a 'network visualization' nowadays. But while Moreno indeed spearheaded the visual display of the network data produced by the sociometric test, his 'mathematical study' was severely deficient in terms of actual mathematical method, which prompted criticism 'for lack of the methodological stringency appropriate to science'[25]. Moreno's work did however succeed in garnering a lot of interest and the journal *Sociometry*, founded in 1937, attracted a number of contributors with mathematical skill, who started to work on the peculiarly rich and complex data produced by an ostensibly simple test. In 1946, Forsyth and Katz made a critical step towards mathematization – the very image of 'methodological stringency' – by proposing to replace the networks drawn in a 'process of trial and error' with a square matrix, in order to 'present sociometric data more objectively'[26]. While the goal was to create a less confusing way of *displaying* a graph[27], the authors also produced a quasi-algorithmic technique to *manipulate* the matrix in a way that subgroups would become visible, simply by reordering rows and columns. This shift in mode of representation opened the door for the application of relatively simple put powerful methods from matrix algebra[28]. Just like statistics provides a number of (standardized) techniques, e.g. to establish a coefficient for the level of correlation between two variables, matrix calculations produce 'views' on sociometric data, structural measures that are essentially 'order from order'[29], transformations of numbers into other numbers. We should not make the mistake, however, to dismiss these calculations as *tautological*, as essentially expressing the same thing as the original data; rather, they are examples for 'mechanical reasoning', transformations that become mediations – ways of exploring the world and of making sense of it in terms that are simultaneously *reductionist* (phenomena are reduced to a point and line model) and endlessly *generative* (the number of possible transformations is infinite). Furthermore, the matrix approach made it possible to connect the work on sociometric data to a second line of development in the mathematization of social structure: during the 1940s, Bavelas[30] had developed an innovative sketch for the mathematical expression of group structures, building on Kurt Lewin's *Principles of Topological Psychology*[31]. Using mostly geometric methods, he introduced such concepts as *distance* and *centrality* into the social network vocabulary. As Festinger argued, the matrix approach was useful here as well, because it could be applied directly to 'such concepts as diameter of [topological structures] and distance from one part of the structure to another'[32] and produce appropriate measures.

According to Barnes and Harary, the 'first indisputable application of graph theory to network analysis did not come until 1953'[33] when Harary and Norman's short volume Graph Theory as a Mathematical Model in Social Science[34] was published. While most of the attempts at constructing mathematical methods for expressing and analyzing sociometric data up to this point had come from sociologists and psychologists with an interest in mathematics – with the notable exception of the statistician Leo Katz –, this was the work of two mathematicians with an interest in the social sciences. Frank Harary in particular established himself as the key actor in this project of 'mathematizing mathematization', even before the publication of his seminal *Graph Theory*[35], which not only standardized concepts, vocabulary, and methods, but also presented them in a pedagogical, applicable way that was crucial for the spread of graph theory.

While the more rigorous mathematical approach did not mean that ties with the social sciences were severed – Harary regularly collaborated with social scientists throughout his long career – there were three important consequences. First, pictorial representations of networks and visual forms of analysis faded into the background. Network diagrams were still used, but rather as teaching aids in textbooks than as research tools in empirical work. Only the spread of graphical user interfaces in the 1990s and the development of layout algorithms based on physics simulations lead to a true renaissance of the now so familiar point and line diagrams. Second, network metrics and methods for mathematical analysis proliferated and became both conceptually and computationally more demanding, sometimes to a point where their empirical applicability became technically forbidding and methodologically questionable, even to scholars sympathetic to the general approach. Third, while exchange between the mathematical methods of graph theory and empirical work in the social sciences has stayed strong over the last decades, the movement towards abstraction and 'purification' implied by mathematization has led to a certain demarcation between the two. On the one side, graph theory has turned more and more into what Latour[36] calls an 'immutable mobile', an inscription or a method that can travel from one setting to another without losing its shape, ready to be *applied*. On the other side, there has been a certain tendency, especially in less 'quantitative' circles of sociology and social theory[37], to use the network model as an ontological concept rather than an analytical one, and to endow it with certain properties, e.g. decentralization, uncontrollability, opposition to hierarchies, and so on. In contrast, the epistemological commitment that underpins the application of graph theory to social phenomena is not to the 'network' as a positive empirical *entity* but, primarily, to empirical and mathematical *concepts* and *methods*. Consider the following quote:

'As we have seen, the basic terms of digraph theory are point and line. Thus, if an appropriate coordination is made so that each entity of an empirical system is identified with a point and each relationship is identified with a line, then for all true statements about structural properties of the obtained digraph there are corresponding true statements about structural properties of the empirical system.'[38]

What takes shape in these lines is a separation between the empirical and the analytical, characteristic to quantitative empirical research, that shifts the full epistemological weight onto the shoulders of *formalization*, that is, onto the 'appropriate coordination' between the two levels. I do not want to contest this separation in philosophical terms here, but rather put forward the much simpler critique that in order to produce a formalization of social phenomena that can be 'mapped' unto the axioms of graph theory, a commitment has to be made, not to the 'network' as an ontological category, but rather to a *theory of the social* that supports and substantiates the formalization process. In Moreno's case, for example, it is a theory of psychological 'attractions and repulsions' between 'social atoms' that *bears* the epistemological weight of formalization in the sense that it *justifies* the mapping of a relationship between two people onto two points and a line by conceiving 'the social' as a primarily dyadic affair. I will come back to this question further down, after having looked at the historic development of a *specific* type of measure that develops against the backdrop I have just outlined: the recursive status index.

2.2. Recursive Status

While over the last decades social theorists[39] have regularly opposed network forms of social organization to hierarchies, the empirical-analytical approach followed by small-group research makes no such opposition; quite to the contrary, graph theory is seen as a means to detect and to measure social hierarchies of different kinds. As mentioned above, despite the definition of sociometry as 'mathematical study', Moreno's mathematical toolkit was rudimentary. But the sociometric test yielded such seemingly simple and highly formalized data that even basic *counting* could lead to curious discoveries, e.g. concerning the distribution of interpersonal choice. Transferring such a 'choice distribution', recovered from the study of the *New York State Training School for Girls*[40], to New York City, Moreno makes an interesting observation that I would like to quote in full here:

> 'For New York, with a population of 7,000,000, the above percentages would be after the 1st choice, 3,200,000 individuals unchosen; after the 2nd choice, 2,100,000 unchosen; after the 3rd choice, 1,400,000 unchosen; after the 4th choice, 1,200,000 unchosen; and after the 5th choice, 1,050,000 unchosen. These calculations suggest that mankind is divided not only into races and nations, religions and states, but into socionomic divisions. There is produced a socionomic hierarchy due to the differences in attraction of particular individuals and groups for other particular individuals and groups.'[41]

The idea that social structure is hierarchic – even in the *absence* of explicit, institutional modes of hierarchical organization – and that a network approach can identify these stratifications is recognizable here, eighty years before a 'new' science of networks[42] began to find power-law distributions in connectivity in (most of) the places it looked. But Moreno's 'socionomic hierarchy' is only the beginning of a long lasting and constant relationship between applied network mathematics and hierarchies of various kinds.

A more formal analysis of differences in *status, authority*, and *influence*, the three terms most frequently – and often interchangeably – used in this context, developed with the matrix approach outlined above. One of the techniques Festinger[43] put forward consisted of simply squaring a matrix (multiplying it by itself), which results in a new matrix where each entry shows the number of two-step connections between two nodes. Cubing the matrix would get the numbers for three step paths, and so on. If applied to data from a sociometric test asking participants to indicate the persons 'influencing' them, this technique would show who 'influences the greatest number of people in less than a specified number of steps'[44]. The question of social *power* becomes a question of *calculation*.

Here, a specific paper, cited in both *PageRank* patents, stands out: in 1953, Leo Katz publishes *A New Status Index Derived From Sociometric Analysis* in the journal *Psychometrika* and introduces the notion of a 'recursive' or 'cumulative' status index. Building on his earlier work on matrix representations of data collected by Moreno[45], Katz proposes a 'new method of computation'[46] for calculating social status from sociometric data. Status measures were already common in studies using the sociometric test, but they were essentially based on simply counting 'votes', as seen in the quote by Moreno above. Katz explicitly *rejects* this method and argues that 'most serious investigators […] have been dissatisfied with the ordinary indices of "status," of the popularity contest type' because this 'balloting' ultimately would not allow to 'pick out the *real* leaders'[47]. His goal was not to measure *popularity* but *social power*, even if the latter term was not explicitly used. Therefore, Katz proposed a new index that takes into account '*who* chooses as well as how many choose'[48], which means that a vote from 'small fry'[49] would simply count less. This shift rests on the idea that status is 'cumulative' in the sense that the topology of a social network expresses a latent 'socionomic hierarchy' in which the status of an individual largely depends on the status of her network neighborhood. Who you are is who you know.

Katz proposes a technically simple but conceptually challenging technique of transforming the initial sociometric 'choice matrix' by first computing all the paths between all the nodes, then attributing a lower weight to longer paths through 'damping', and finally calculating a metric for every node based on the combined weight of their connections to all other nodes[50]. Similar measures to this one, today known as *Katz centrality*, abound in the following decades and are also used to other ends than status measuring, e.g. for the identification of cliques by Hubbell in 1965[51]. The contribution by the mathematical sociologist Phillip Bonacich in 1972[52] is worth noting, because it consists of a metric based on the *eigenvectors* of a matrix, which means that not *all* paths between nodes are taken into account, but only the *shortest* paths, therefore making the metric less dependent on *local* properties of networks[53] and consequently more difficult to 'cheat' by adding nodes and connections locally. This resistance to what could be called 'link spam' is probably the reason why *PageRank* is a variant of eigenvector centrality rather than Katz centrality, although references to Bonacich's work are conspicuously absent from the *PageRank* patents.

A second arena where concrete techniques for the mathematical exploration and measurement of networks are pioneered – and actually applied to decision-making – is citation analysis. In 1963 the *Institute for Scientific Information*, founded by Eugene Garfield, publishes the first edition of the *Science Citation Index* (SCI), an index of citations, manually extracted but sorted by computer, from 613 journal volumes published in 1961. With the first edition storing already 1.4 million citations on magnetic tape, this index is perhaps the first 'big data' file available in the social sciences, and over the following years a significant number of researchers participate in analyzing it with various computational methods[54]. Despite Eugene Garfield's intention[55] to promote the SCI first and foremost as an 'association-of-ideas index', a tool for *finding* scientific literature, it quickly becomes obvious that a series of evaluative metrics could be derived from it without much effort. Thus, in 1972, Garfield presents a fleshed-out version of a concept he had initially introduced as a tool for the historiographical study of science, the (in)famous 'impact factor', now adding that '[p]erhaps the most important application of citation analysis is in studies of science policy and research evaluation'[56]. When taking into account that the first attempt at 'citation ranking' by Gross and Gross in 1927[57] was designed to help university libraries in deciding which chemical journals they should subscribe to, it becomes clear that any evaluative metric would instantly face considerable normative and institutional *entanglement*, much more so than in the case of the (mostly) descriptive sociometry. From a mathematical standpoint, the impact factor, calculated for scientific journals, is an extremely basic measure: it takes the number of citations the last two volumes of a journal received the following year and divides it by the number of articles published in these two volumes. This second aspect leads to an important question – and certainly a central reason why literature from the citation analysis field is cited in the *PageRank* patents – that is not addressed in sociometry, the problem of *size*. If a journal publishes a large number of articles it is well positioned to receive more citations than a journal publishing less; purely counting citations without taking *volume* into account would therefore be misleading and this is why the impact factor divides by the number of articles published[58].

A paper by Pinski and Narin, published in 1976, pushes things significantly further by pointing out two problems with Garfield's measure. First, citations have equal value in the impact factor scheme, although 'it seems more reasonable to give higher weight to a citation from a prestigious journal than to a citation from a peripheral one'[59]. Pinski and Narin therefore propose a *recursive* index for importance, based on the same eigenvector calculations that we found in Bonacich's work, although the authors were apparently not aware of the work done in sociometry. Second, Pinski and Narin argue that the impact factor attributes disproportional importance to review journals, and 'can therefore not be used to establish a "pecking order" for journal prestige'[60]. Again, they propose a 'solution' by including an 'input-output' measure into their calculations[61] where the weight (W) of incoming citations (C) is divided by the number of outgoing citations (S):

$$W_i = \sum_{k=1}^{n} \frac{W_k C_{ki}}{S_i}$$

If we were to introduce a 'damping factor' into this model, such as Katz's connection weight reduction scheme, in order to attenuate the 'free flow of citations in the referencing marketplace'[62], we would essentially end up with *PageRank*. While Pinski and Narin do not quote the work of economist Wassily Leontief, unlike Hubbell[63] who acknowledges this source of inspiration in his sociometric clique identification scheme, there is reason to believe that their version of an input-output model was equally inspired by economic thought, in particular if we take into account that Pinski and Narin's metric was intended for a 'funding agency with its need to allocate scarce resources'[64].

This appearance of a conceptual horizon based on economic concepts can be observed in the sociometric field as well and merits a moment of attention.

2.3. From Sociometry to Social Exchange Theory

While the sociometric test was generally lauded as a tool for producing interesting data, Moreno's socio-psychological theory of 'attraction and repulsion' between 'social atoms'[65] was based on highly contested assumptions and his goal to develop a 'technique of freedom' to reorganize society by training individuals to transcend their social prejudices in order to liberate the forces of 'spontaneous attraction', did not go well with the sober and pragmatic mindset of most empirical sociologists. The sociometric papers working with Moreno's data therefore generally subscribed to a vague version of the same atomistic and dyadic view of society, which enabled and justified the 'point and line' formalization needed to apply graph theoretical methods, but shunned the deeper aspects of the theoretical horizon behind it. Concepts like influence, status, importance, prestige, and so on fill the void, but they are defined in experimental terms reminiscent of Paul Lazarsfeld work[66] rather than as parts of robust theoretical frameworks. Influence, for example, if defined at all, is generally conceived as the capacity of an individual to change somebody else's 'opinion', as measured e.g. by a vote in a political election. Even Kurt Lewin explicitly acknowledges that the goal of mathematization overruled other considerations:

> 'We have tried to avoid developing elaborate "models"; instead we have tried to represent the dynamic relations between the psychological facts by mathematical constructs at a sufficient level of generality.'[67]

And Kurt Lewin was certainly the sociometric scholar with the most substantial theoretical ambitions: in a chapter entitled *The Light that Failed*, Nicholas Mullins[68] squarely attributed the demise of 'small-group theory' to the intellectual and theoretical vacuum left by Lewin's death in 1947. In his famous piece *The Strength of Weak Ties*, Granovetter wonders likewise why '[s]ociometry, the precursor of network analysis, has always been curiously peripheral – invisible, really – in sociological theory'[69] and attributes this not only to the difficulty of adapting methods to larger group sizes, but also to a lack of 'theoretical detail', combined with 'a level of technical complexity appropriate to such forbidding sources as the *Bulletin of Mathematical Biophysics*'[70]. But an analysis of Granovetter's own work indicates where the field turns to find a new and more concise conceptual horizon: to social exchange theory, an approach that 'might be described, for simplicity, as the economic analysis of non economic social situations'[71]. Here, power is no longer defined in vague psychological terms but in a sober, 'hard-knuckled' way as access to resources for economic advancement – essentially to *information* on jobs, prices, opportunities, and so on – which can be modeled in terms of graph theory just as easily as attraction and repulsion before that. Having been turned into immutable mobiles by Harary and others, the mathematical concepts, methods, and metrics travel effortlessly into the new conceptual space[72] that, notably, does *not* break with an atomistic [73] and dyadic conception of social relations. What in Moreno's view had to be *produced* through spontaneity training, namely the free association of individuals, now conveniently becomes a basic property of individuals conceived as free actors engaged in the 'rational choice' of utility maximization. While Mayer[74] rightfully points to the sociometric heritage – and its closeness with cybernetics and operations research – of contemporary search engine metrics, I would argue that the *economic* trajectory represents an equally important lineage. First by supplying a rich new space of conceptual inspiration, as I have argued above for the case of the input-output model, and second by providing a new 'home' for the mathematical metrics and methods pioneered in sociometry.

It is a core paradox of mathematical network analysis that it provides powerful tools for examining concrete power relations empirically, which holds considerable critical potential, and, at the same time, that it allows using the measured hierarchies as operative mechanisms to allocate resources such as visibility or research funding. What all of the strands I have presented here furnish – and this is fundamental for any form of *normative* and *operational* application of evaluative metrics, whether in citation ranking or in Web search – is a narrative that sustains what I would like to call the 'innocence of the link': whether it is spontaneous attraction, rational choice or simply an 'inspirational' account of scientific citation, the application of the metrics to actual ranking, with concrete and tangible consequences, can only be justified if the link is kept reasonably pure. In this vision, the main 'enemy' is therefore the *deceitful linker*, whether they come in form of scientific citation cartels or their contemporary cousins, link farms. It is not surprising that a central argument against citation analysis as a means for research evaluation builds on a critique of actual citation practices.

In this section, I have tried to sketch a historical background for certain core ideas present in *PageRank* in order to provide the 'material' – the *resources*[75] – to contextualize this evaluative metric *conceptually*, in the sense that a critical reading of formulas or source code not only requires a capacity to understand technical languages but also 'interpretive ammunition' to *refill* computational artifacts that have most often been cleansed from easily accessible markers of context and origin. The historical approach is thus *one* way of bringing light into a black box. If epistemological and methodological concepts such as the network model suggest 'a reading of reality'[76], the historical approach can provide the background for a conceptual analysis of the logic this reading implies, of the motivations behind it, and of the operations that go into actually *making it work*. Such an analysis could seek to identify general characteristics of the network approach, for example the tendency to highlight 'the synchronic dispersal over the diachronic unfolding'[77]; it could situate it in relation to certain theories of the social and to specific mathematical methods, as I have tried to do in this section; but it could also consist of further closing in on actual technical artifacts, which is the objective of the next section.

3. Two Moments of Commitment

In this section, I will develop two axes of interpretation for the *PageRank* model, a comparative approach and a 'close reading' of a particular parameter, both organized around the question of how an understanding of metrics as *descriptions* can help us in dealing, analytically, with their becoming operative *prescriptions*.

While citation metrics like the impact factor have quickly crossed the threshold between 'is' and 'ought', their use in 'science policy and research evaluation'[78] is hardly automated. Much like other statistical surveys or indicators before them, they have come to play an essential role in bureaucracies where ideals of 'evidence-based' governance privilege mathematical means of decision-making, which are seen as providing 'rules of discourse so constraining that the desires and biases of individuals are screened out', therefore guaranteeing 'the rule of law, not of men'[79]. But even as integral parts of bureaucratic processes, these metrics retain both a certain visibility, which makes them amendable to critique, and margins of discretion in terms of how they are used. The rendering of graph theoretical methods in software is certainly significant on the *descriptive* side of things – just consider how SPSS plays as role in orienting research practices in the social sciences[80] – but *prescriptive* power is particularly enhanced when metrics are *sunk into the system* and become both invisible and *inevitable*, in the sense that margins of discretion are defined by the system itself. When using the Google search engine, neither the ranking principles, nor their parameters are amendable to user intervention.

In 1992 Botafogo, Rivlin, and Shneiderman[81] first explored the transposition – or 'sinking' – of sociometric methods into a *navigational* software environment as a means to solve the 'lost in hyperspace' problem. Building on a paper by Harary, named *Status and Contrastatus*[82], they formalized a hypertext as a directed graph and began to calculate metrics for the purpose of 'recovering lost hierarchies and finding new ones'[83]. The goal was not to find the most important document nodes for document retrieval, but rather to assist hypertext authors in *designing* the structure of their text networks more explicitly – 'remember that structure does reflect semantic information'[84]. Structural hierarchies were seen as helpful navigational devices and status metrics should make it easier to *build* them. Interestingly though, Botafogo, Rivlin, and Shneiderman underscored the idea that software would make it easy to implement different network metrics and thereby provide not just one 'view' on the hypertext, but many different ones, granting the 'ability to view knowledge from different perspectives'[85].

Taking all of the work in sociometry, citation analysis, and hypertext navigation together, one could argue that all the 'ingredients' for *PageRank* were available from the middle of the 1990s, and that all one had to do was to combine them. It is not the purpose of this paper, however, to judge the 'originality' of Page and Brin's work – a somewhat foolish undertaking in the light of actor-network theory – but rather to show that network metrics were developed in different contexts and with different purposes in mind, and that they were indeed made operational in different ways and at different levels of performativity. This is where it is useful to connect back to Foucault's notion of the archive, understood not simply as a coherent *paradigm* in Kuhn's sense, but rather as a wider 'conceptual space'[86] that allows for significant variation and dissension despite the agreement on certain fundamentals. This means, however, that an overarching critique of the reductionist and atomistic perspective espoused by the network model runs the risk of loosing sight of the considerable space of choice available on the inside of the conceptual space examined here. In the remainder of this paper I would therefore like to argue that we need to be attentive not only to moments of totalization, homogenization, and closure but also to vectors of variation,

heterogeneity, and fermentation, in terms of both concepts and politics. If sociometric measurement means that 'authority no longer rests in the [social] relationships and instead migrates towards the measuring instrument'[87], the question remains how we can better understand the way authority is configured, concretely, once it has migrated. I believe that engaging *PageRank* as a *computational model* can bring us closer to an answer; but it requires that we do more than capture its *spirit*; we have to attempt to get to its *technical logic*. By comparing it to one of its many siblings, Jon Kleinberg's HITS algorithm, I will develop a first sketch of what such an attempt could look like.

3.1. Flatlands and Two Ways to Make Hills

It is rather remarkable that, despite being initially conceived as a technique for *finding* scientific literature, citation analysis was not combined with existing techniques from the field of *information retrieval* (IR) – a term coined by Calvin Mooers in 1950 and a blossoming field of research and product development since the 1960s – before the 1990s. It is certainly safe to argue that the Web, this gigantic and unruly network of documents, played a significant role as catalyst in bringing together fields and methods that were largely separate until that point. While I have only followed one of the two technological strands here, the Google search engine is of course both a link analysis and an IR system. At least conceptually, the two components can indeed be analyzed separately and there is no reason to believe that Google's information retrieval engine – the system that looks for the presence of query terms in a document corpus and scores results based on a series of *internal* lexical and structural metrics – was significantly different from the already well established models of the time. Put simply, the shift in perspective did not concern retrieval techniques as such, but rather the way the *corpus* was conceptualized. Attributing an 'importance' score[88] to every document, 'independent of the page content'[89], fundamentally means giving up the idea of a 'flat corpus'[90] of documents where *relevance* is only determined in relation to a query. *PageRank* is the mechanism by which the Web is no longer treated exclusively as a document repository, but additionally as a *social system*. Every document in the corpus is already projected as a member of a stratified network/society before the searching even begins. The classic IR idea of *relevance* – always conceived in relation to a specific 'informational need' – is complemented by the sociometric concepts of *status* and *authority*. More precisely, because linking is framed as the rational attribution of importance, the social system can be seen as a legitimate 'source' of a singular and universal understanding of authority. But is this an inevitable consequence of link analysis and sociometric measurement? The comparison with HITS[91] is instructive in this respect.

Just like *PageRank*, HITS (Hyperlink-Induced Topic Search) is an eigenvector based evaluative metric that was introduced at nearly the same time as its famous cousin. Its goal to formulate 'a notion of authority' by processing the 'considerable amount of latent human judgment' encoded in hyperlinks, in order to solve the 'abundance problem'[92] faced by Web users, is near identical as well. But there are two major differences: First, HITS actually provides two eigenvector metrics, one for 'hubness' and another for 'authority'. Building on the idea 'that good hubs point to good authorities and good authorities are pointed to by good hubs'[93], Kleinberg makes a conceptual differentiation, entirely based on structural properties of the graph, that results in a typology of nodes and, potentially, the possibility for users to specify the type they are looking for. While metrics indeed totalize and *commensurate*[94], there *are* technical means to introduce variation and multiple perspectives. Second, HITS inverts the temporal sequence of authority ranking and document retrieval. Rather than calculating a *universal* or *a priori* landscape of authority, documents matching a query are retrieved first and authority is calculated second – thus based on the link structure in the result set only and not the full corpus. This means that in the HITS perspective, authority is dependent on a domain – expressed through a query – and a page that receives a high authority score for one search term may well get a much lower score for another. While this obviously does not abolish the notion of authority by any means, the concept is reconfigured, contextualized, and 'contained' in the sense that HITS is less subject to what Barry and Brown[95] euphemistically call 'topic drift', i.e. the dominance of high-authority sites over a wide range of topics.

Software, however, does not exist in a disembodied state where ideas simply compete on the merit of their outputs, and while we may want to argue that the HITS perspective holds significant advantages over *PageRank*, these arguments clash with the fact that the computational requirements for HITS are much, much higher than for *PageRank*. While the latter can recalculate its universal landscape of authority at a set interval – before it fell silent on the topic, Google indicated a number of about once a month – the former would have to make an authority computation *for every single search request*. Certainly, the scores for the most popular queries could be calculated in advance, but a large disadvantage persists when it comes to both speed and cost. But even when looking *inside* the PageRank formula, we find space for variation and choice. In the next section, I will show how a single parameter encodes a significant theoretical, and consequently operational, commitment.

3.2. Objet petit α

Just like Katz's 'status index'[96] and Bonacich's 'power centrality'[97], the *PageRank* model, which 'recursively propagate[s] weights through the link structure of the Web'[98] includes 'a damping factor that limits the extent to which a document's rank can be inherited by children documents'[99]. The rationale put forward to explain rank 'decay'[100] in *PageRank* is twofold: first, the actual problem in the context of a Web search engine is that without damping, a 'rank sink' in the form of a closed loop between two or more pages could excessively accumulate rank at each pass of the iterative approach to computing *PageRank*, which is the preferred method of calculation; damping eliminates those undesired artifacts. Second, an 'intuitive basis'[101] in the form of the model of a 'random surfer' is introduced, where the damping factor α 'models the fact that users typically jump to a different place in the web after following a few links'[102]. While *PageRank* is presented as 'an attempt to see how good an approximation to "importance" can be obtained just from the link structure'[103], neither the 'rank sink' nor the 'random walk' justification sufficiently account for damping in respect to the rationale of *importance*. The first element is a technical problem arising from the particular computational method used and the second does not elaborate on how a model of random user behavior relates to the calculation of an importance score. While *PageRank* can indeed be read as a scoring of the 'reachability' of a node in a graph, the transfer from reachability to importance remains unexplained here. We can, however, develop a line of interpretation by considering the explanations Katz and Bonacich provide for the inclusion of a damping factor into their own metrics.

Katz introduces his α factor as a means to adapt the calculation of his 'status index' to 'both the group and the context'[104] and provides an expressive example from the domain of information diffusion:

> 'For example, the information that the new high-school principal is unmarried and handsome might occasion a violent reaction in a ladies' garden club and hardly a ripple of interest in a luncheon group of the local chamber of commerce. On the other hand, the luncheon group might be anything but apathetic in its response to information concerning a fractional change in credit buying restrictions announced by the federal government.'[105]

What Katz seeks to encode into α is an estimate of the 'probability of effectiveness of a single link'[106] and because already small variations in the factor can lead to significant variations in the status index, the *commitment* to an *appreciation* of the 'conductivity' of a network – lets not forget that α is a manually set constant – has substantial consequences on the calculated hierarchy. Similarly, Bianchini et al. present α in *PageRank* as the factor deciding *how far* the structure of a network should influence the status of a node:

> 'PageRank deeply depends on parameter [α]. When [α] is small, only the nearest relatives give an actual contribution to [PR], whereas, as [α] → 1, far ancestors are also significant for the evaluation of [PR].'[107]

If α is small, ranking actually approaches the 'balloting' Katz criticized as a 'dissatisfying' measure of status, because status no longer propagates and the simple number of 'votes' a person receives becomes the determining factor.

The entanglements of this modulation become clearer when examining the arguments behind Bonacich's decision[108] to add a damping factor to his measure of 'power centrality' (PC). In order to account for the results of an exchange theory based study by Cook et al.[109], which had indicated a disjunction between centrality and power in a bargaining network, Bonacich proposes a parameter β, which makes it possible to adjust the depth of influence in a fashion very similar to α by defining 'a radius within which power or centrality is being assessed'[110]. What makes Bonacich's piece so interesting is his reflection on how β becomes a way to implement varying conceptions of what constitutes power in *different* social networks. The full quotation is instructive:

> 'To some, the measure [PC] may seem hopelessly ambiguous; [PC] can give radically different rankings on centrality, depending on the

value of β. However, the measure accentuates an inherent ambiguity in the concept of centrality. There are different types of centrality, depending on the degrees to which local and global structures should be weighted in a particular study and whether that weight should be positive or negative. When communication is typically over long distances, position in the global structure should count more than when all communication is local. In an organized hierarchy in which power is transitive, the power of those one has power over should be weighted more highly in determining overall power than when all relations are dyadic.'[111]

Bonacich's β – and this is the difference with α – is designed to work with a *negative* value as well, so that connections to high status nodes are *detrimental* to rank. This allows the model to account for bargaining networks, where exchange theory considers it beneficial to be connected to a large number of 'low status' nodes that can be more easily exploited because they lack information, e.g. about prices. In this situation, 'power comes from being connected to those who are powerless'[112]. Because β and α are set manually, they express a true commitment to a theory about the real-world properties of the network in question.

Applied to *PageRank*, two very different notions of status, authority, and importance emerge: with α → 0, the Web would indeed be conceived as a 'one person, one vote' type democracy[113] and collecting inlinks, no matter from where, would be the principal means to gain rank; in a α → 1 setting, the Web becomes a very different place – opaque, complex, and potentially Machiavellian, where success comes either from patronage by the already powerful or the patient and costly construction of a large and spread-out strategic network. While we have to keep in mind that this is a scale and not a binary decision, the first vision puts an emphasis on the local properties of the graph and makes it therefore possible for local changes, if significant in number, to affect the overall ranking of the system; the second however builds on the global structure of the graph and local changes therefore have negligible effects on overall rankings. If we keep with the image of a social system, the respective models of transformation are *populist/revolutionary* and *conservative/reformatory*. But while purely descriptive metrics make an *epistemological* commitment to a theory of power by using it to *represent* social status, the prescriptive mode of *PageRank* makes an *operational* commitment to *reproduce*, in very practical terms, the hierarchy of social status it detects. The value Google usually communicated for α was 0.85[114] and at this level, the often-heard interpretation of *PageRank* as essentially an expression of *popularity*[115] misses the target. While sheer numbers certainly do count, the idea that PageRank interprets the Web as 'a gigantic popularity contest'[116] is simply too imprecise.

If we consider the Google search engine as a central site of power negotiation and arbitrage for the Web, an analysis focusing on *PageRank* – which in practice is certainly not enough – would have to conclude that its authority ranking mechanism applies, in a universal fashion, a largely conservative[117] vision of society to the document graph in order to 'pick out the *real* leaders'[118] and distribute visibility to them. Rather than showing us the *popular* it shows us the *authoritative*, or, to connect back to citation analysis, the *canonical*[119]. If we consider the link indeed as 'innocent', as a valid indicator of disinterested human judgment, *PageRank* shows us a *meritocracy*; if we take the link to be fully caught up in economic forces however, we receive the map of a *plutocracy*. The search engine as a visibility engine subjects both to the self-reinforcing dynamic of *cumulative advantage*.
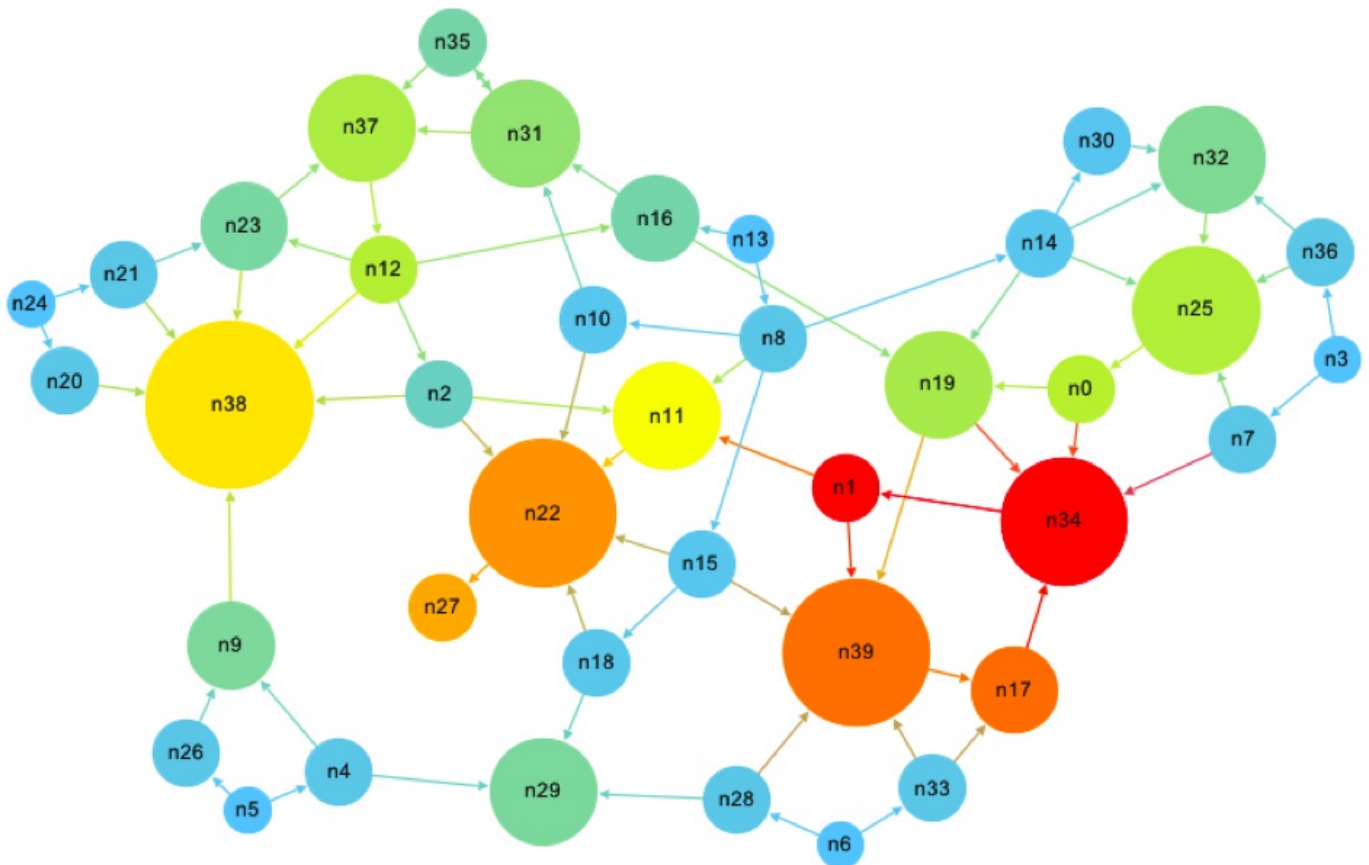

Fig. 1: A network generated randomly with the gephi graph analysis toolkit. Node sizes represent the number of inlinks and color represents PageRank (α=0.85) via a heat scale (blue => yellow => red).

Label	PR α=0.85	PR α=0.7	PR α=0.55	PR α=0.4	In-Degree	Out-Degree	Degree
n34	0.0944	0.0743	0.0584	0.0460	4	1	5
n1	0.0867	0.0617	0.0450	0.0345	1	2	3
n17	0.0668	0.0521	0.0423	0.0355	2	1	3
n39	0.0663	0.0541	0.0453	0.0388	5	1	6
n22	0.0619	0.0506	0.0441	0.0393	5	1	6
n27	0.0591	0.0451	0.0371	0.0318	1	0	1
n38	0.0522	0.0561	0.0542	0.0486	6	0	6
n11	0.0492	0.0372	0.0306	0.0274	3	1	4

Fig. 2: A table of eight nodes from the same network as above, selected and ordered based on the highest values for PageRank (α=0.85). Four PageRank rankings were calculated in 0.15 intervals. Red marks the highest value per column, yellow the second highest, and blue the third. Note how n1 with just a single inlink slips with lower values for α and n38 with six inlinks rises to the top.

Although computer scientists have studied the effects of α extensively[120], Fig. 1 and Fig. 2 show the effects of changes in the value for α on a simple and actually rather 'egalitarian' network in a hopefully more accessible form than the papers written by mathematics specialists. While we cannot make any far-reaching generalizations based on a small experiment, it is in line with the finding by the cited authors that small variations in α can lead to significant changes in ranking. It also confirms that at high values for α, the source of inlinks is much more relevant than simply their number; a single link from a page with a high rank can be extremely valuable. Although this is only a sketch, an approach by means of *tinkering* with an algorithm should be considered as a way to study it as a *model* in its 'natural habitat': eating numbers and spewing out different numbers at the rear end. But again, considerations of computational cost add an interesting caveat. While the level set for α has important conceptual – and potentially political – consequences, it also implies a cost factor because a higher value means that the calculation will require a higher number of iterations to converge.[121] Langville and Meyer capture the dilemma:

> 'Consequently, Google engineers are forced to perform a delicate balancing act. The smaller α is, the faster the convergence, but the smaller α is, the less the true hyperlink structure of the web is used to determine webpage importance. And slightly different values for α can produce very different PageRanks.'[122]

This means that a local theory of power is computationally and, because computation costs money, economically *cheaper* than a theory where status is an effect of more global structural properties. In this case, the menace from spam apparently looked sufficiently substantial to convince Google that a larger radius of influence – and consequently a reduced capacity for local change to affect global structure – was well worth the additional cost.

4. Conclusions

This paper set out to examine what is *in* the *PageRank* algorithm and it proceeded in two ways: first, by situating this evaluative metric in a larger *archive* of ideas, concepts, theories, and methods that developed around sociometry, citation analysis, social exchange theory, and hypertext navigation. This conceptual space constitutes, at least to a certain extent, the place of emergence from where *PageRank* became 'sayable'[123]. Second, by comparing the algorithm to a close cousin and by examining a particular parameter in the model, I have tried to show that the concrete artifact – the model, not the search engine – does not simply follow from the historical *a priori*. There are thousands of graph theoretical algorithms documented and hundreds of variants of the recursive status metrics I have focused on. Are they all based on a reductionist and atomistic vision of social relations? Quite possibly. Do they all encode the same logic, produce the same rankings, and commit to the same politics? Certainly not. There *is* variation and it is *significant*.

This paper has dedicated much attention to mathematics and relatively little to software as execution of instruction written in a programming language. Does this mean that I am advocating that we should analyze software as a 'special case' of mathematics, as the mechanical equivalent of computational theory? Not at all. Most programmers rarely have to venture beyond (junior) high school mathematics and spend most of their time programming interface states and specifying data logistics. What is *in* most lines of code are things that have been done a million times by a million people, and a researcher can get a sufficient idea of how many systems work by clicking around the interface in a somewhat systematic fashion. In other cases, however, it can certainly be useful to read the source code of an application to understand what it does and how it does it. For yet another set of cases, a written documentation, a class diagram, a database schema, and so forth can be decisive or even sufficient. And then there are objects like *PageRank*, where, in a sense, the mathematical notation 'says it all' and implementations only vary in terms of computational requirements, but we *still* have to brood over it for hours or days, play with it, test it with different data sets, read about it, look into where it comes from, who build it, for what purpose, and so forth, to even begin to understand what it actually *does* in a non-trivial sense. It is hard to imagine what would be required to come to a detailed and robust understanding of the full Google search engine and not just a small piece of it.

What we need in order to analyze an algorithm or a piece of software will vary from case to case, but most of the time, a wide array of *resources*[124] will have to be mobilized, an array than can span the space between a reeducation camp for girls in the 1930s and a small parameter bearing on a single line of code and still be far from exhaustive without actual users, actual data, and actual societies around them.

So what is *in* the algorithms producing the evaluative metrics discussed in this text? I would argue that they encode ways of putting things into relation that (can) fundamentally reconfigure how power is constituted, how it operates, and how is may be negotiated. By blending *description* and *prescription* in a particularly pervasive fashion, software introduces a set of new techniques into the arsenal of control, techniques that make commitments to certain conceptions of the social and, by becoming operative machinery, produce specific social consequences themselves. But this transposition into the world of algorithms not only modifies the materiality and performativity of these techniques and often renders them near invisible; paradoxically it also *opens them up* to new ways of pondering, criticizing, and – maybe – changing them. We risk missing a genuinely political moment if we lose sight of how software can sometimes make it astonishingly *easy* to do things differently.

5. Bibliography

Andreessen, Marc. 2011. 'Why Software Is Eating The World.' *The Wall Street Journal*, August 20. Accessed April 18, 2012. http://online.wsj.com/article/SB10001424053111903480904576512250915629460.html.

Barnes, John A. and Frank Harary. (1983). 'Graph theory in network analysis.' *Social Networks* 5:235-244.

Bavelas, Alex. 1948. 'A Mathematical Model for Groups Structures.' *Human Organization* 7:16-30.

Bechetti, Luca and Carlos Castillo. 2006. 'The Distribution of PageRank Follows a Power-Law only for Particular Values of the Damping Factor.' In *Proceedings of the 15th international conference on World Wide Web*. New York: ACM.

Berry, Michael W., and Murray Browne. 2005. *Understanding Search Engines: Mathematical Modeling and Text Retrieval. 2nd Edition*. Philadelphia: Society for Industrial and Applied Mathematics.

Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven CT: Yale University Press.

Berry, David M. 2008. 'The poverty of networks.' *Theory, Culture and Society* 25:364-372.

Bianchini, Monica, Marco Gori and Franco Scarselli. 2005. 'Inside PageRank.' *ACM Transactions on Internet Technology* 5:92–128.

Bonacich, Phillip. 1972. 'Factoring and weighting approaches to status scores and clique identification.' *Journal of Mathematical Sociology* 2:113-120.

Bonacich, Phillip. 1987. 'Power and Centrality: A Family of Measures.' *American Journal of Sociology* 92:1170-1182.

Botafogo, Rodrigo A., Ehud Rivlin, and Ben Shneiderman. 1992. 'Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics.' *ACM Transactions on Information Systems* 10: 142-180.

Brin, Sergey and Lawrence Page. 1998. 'The Anatomy of a Large-Scale Hypertextual Web Search Engine.' *Computer Networks and ISDN Systems*

30:107-117.

Castells, Manuel. 1996. *The Information Age: Economy, Society and Culture. Volume 1: The Rise of the Network Society*. Cambridge MA, Oxford: Blackwell.

Chakrabarti, Soumen. 2003. *Mining the Web: Discovering Knowledge from Hypertext Data*. San Francisco: Morgan-Kauffman.

Chun, Hui Kyong Chun. 2011. *Programmed Visions: Software and Memory.* Cambridge MA, London: The MIT Press.

Cook, K. S., R. M. Emerson, M. R. Gilmore, and T. Yamagishi. 1983. 'The Distribution of Power in Exchange Networks: Theory and Experimental Results.' *American Journal of Sociology* 89:275-305.

Diaz, Alejandro. 2008. 'Through the Google Goggles: Sociopolitical Bias in Search Engine Design.' In *Web Search: Multidisciplinary Perspectives*, edited by Amanda Spink and Michael Zimmer, 11-34. Berlin: Springer.

Emerson, Richard M. 1976. 'Social Exchange Theory.' *Annual Review of Sociology* 2:335-62.

Erdös, Paul. 1977. 'A Note of Welcome.' *Journal of Graph Theory* 1:3.

Espeland, Wendy, and Mitchell L. Stevens. 1998. 'Commensuration as a Social Process.' *Annual Review of Sociology* 24: 313-343.

Festinger, Leon. 1949. 'The Analysis of Sociograms using Matrix Algebra.' Human Relations 2:153-158.

Forsyth, Elaine, and Leo Katz. 1946. 'A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report.' *Sociometry* 9:340-347.

Foucault, Michel. 1969. *L'Archéologie du savoir*. Paris: Gallimard.

Fuller, Matthew. 2008. *Software Studies: A Lexicon*. Cambridge MA, London: The MIT Press.

Galison, Peter. 1997. *Image & logic: A material culture of microphysics*. Chicago: The University of Chicago Press.

Garfield, Eugene. 1955. 'Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas.' *Science* 122:108-111.

Garfield, Eugene. 1972. 'Citation Analysis as a Tool in Journal Evaluation.' *Science* 178: 471-479.

Gitlin, Todd. 1978. 'Media Sociology. The dominant paradigm.' *Theory and Society* 6:205-253.

Golumbia, David. 2009. *The Cultural Logic of Computation. Cambridge*, London: Harvard University Press.

Granovetter, Mark S. 1973. 'The Strength of Weak Ties.' *The American Journal of Sociology* 78:1360-1380.

Gross, P. L. K., and E. M. Gross. 1927. 'College Libraries and Chemical Education.' *Science* 66:385-389.

Hanneman, Robert A. and Mark Riddle. 2005. *Introduction to social network methods*. Riverside: University of California, Riverside. http://faculty.ucr.edu/~hanneman/

Harary, Frank, and Robert Z. Norman. 1953. *Graph Theory as a Mathematical Model in Social Science*. Ann Arbor: University of Michigan.

Harary, Frank. 1959. 'Status and Contrastatus.' *Sociometry* 22:23-43.

Harary, Frank. 1969. *Graph Theory*. Reading, Menlo Park, London, Don Mills: Addison-Wesley.

Harary, Frank, Robert Z. Norman, and Dorwin Cartwright. 1965. *Structural Models: An Introduction to the Theory of Directed Graphs.* New York, London, Sydney: John Wiley & Sons.

Harary, Frank. 1983. 'Conditional Connectivity.' *Networks* 13:347-357.

Hubbell, Charles H. 1965. 'An Input-Output Approach to Clique Identification.' *Sociometry* 28:377-399.

Introna, Lucas D. 2007. 'Maintaining the Reversibility of Foldings: Making the Ethics (Politics) of Information Technology Visible.' *Ethics and Information Technology* 9:11–25.

Katz, Leo. 1953. 'A New Status Index Derived from Sociometric Analysis.' *Psychometrika* 18:39-43.

Kleinberg, Jon. 1999. 'Authoritative sources in a hyperlinked environment.' *Journal of the ACM* 46:604-632.

Knuth, Donald. 1968. *The Art of Computer Programming. Volume 1: Fundamental Algorithms*. Reading, Menlo Park, London, Don Mills: Addison-Wesley.

König, Dénes. 1936. *Theorie der endlichen und unendlichen Graphen*. New York: Chelsea Publishing Company.

Langville, Amy N. and Carl D. Meyer. 2004. 'Deeper Inside PageRank.' *Internet Mathematics* 1:335-380.

Langville, Amy N. and Carl D. Meyer. 2005. 'A Survey of Eigenvector Methods for Web Information Retrieval.' *SIAM Review* 47:135-161.

Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society.* Cambridge, MA: Harvard University Press.

Lewin, Kurt. 1936. *Principles of Topological Psychology*. New York: McGraw-Hill.

Lewin, Kurt. 1964. *Field Theory in Social Science: Selected Theoretical Papers.* New York: Harper & Row.

Lewis, David, and Andrew J. Weigert. 1985. 'Social Atomism, Holism, and Trust.' *The Sociological Quarterly* 26:455-471.

Manin, Bernard. 1997. *The Principles of Representative Government*. Cambridge: Cambridge University Press.

Manovich, Lev. 2008. *Software Takes Command*. Accessed April 18, 2012. http://lab.softwarestudies.com/2008/11/softbook.html

Mayer, Katja. 2009. 'On the Sociometry of Search Engines: A Historical Review of Methods.' In *Deep Search: The Politics of Search beyond Google* edited by Konrad Becker and Felix Stalder, 54-72. New Jersey: Transaction Publishers.

Moreno, Jacob L. 1934. *Who Shall Survive? A New Approach to the Problem of Human Interrelation*. Washington: Nervous and Mental Disease Publishing.

Mullins, Nicholas C. 1973. *Theories and theory groups in contemporary American sociology*. New York: Harper & Row.

Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. 'The PageRank Citation Ranking: Bringing Order to the Web.' Technical Report, Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/

Page, Lawrence. 2001. Method for Node Ranking in a Hyperlinked Database. U.S. Patent 6,285,999, filed January 9, 1998, and issued September 4, 2001

Page, Lawrence. 2004. Method for Scoring Documents in a Linked Database. U.S. Patent 6,799,176, filed July 6, 2001, and issued September 28, 2004

Pinski, Gabriel and Francis Narin. 1976. 'Citation influence for journal aggregates of scientific publications.' *Information Processing and Management* 12: 297–312.

Porter, Theodore. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life.* Princeton, NJ: Princeton University Press.

Rieder, Bernhard. 2009. 'Democratizing Search? From Critique to Society-oriented Design.' In *Deep Search: The Politics of Search beyond Google* edited by Konrad Becker and Felix Stalder, 133-151. New Jersey: Transaction Publishers.

Roush, Wade. 2004. 'Search Beyond Google.' *Technology Review*, March. Accessed April 19, 2012. http://www.technologyreview.com/articles/04/03/roush0304.2asp

Stigler, Steven M. 1999. *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press.

Uprichard, Emma, Roger Burrows, and David Byrne. 2008. 'SPSS as an 'inscription device': from causality to description?' *The Sociological Review* 56:606-622.

Watts, Duncan J. 2004. 'The 'New' Science of Networks.' *Annual Review of Sociology* 30:243-270.

Zimmer, Michael. 2010. 'Web Search Studies: Multidisciplinary Perspectives on Web Search Engines.' In *International Handbook of Internet Research* edited by Jeremy Hunsinger, Matthew Allen, and Lisbeth Klastrup, 507-522. Dordrecht: Springer.

1. Knuth, Donald. 1968. *The Art of Computer Programming. Volume 1: Fundamental Algorithms.* Reading, Menlo Park, London, Don Mills: Addison-Wesley, 4. ↵
2. Knuth 1968, 3. ↵
3. Andreessen, Marc. 2011. 'Why Software Is Eating The World.' *The Wall Street Journal*, August 20. Accessed April 18, 2012. http://online.wsj.com/article/SB10001424053111903480904576512250915629460.html. ↵
4. Fuller, Matthew. 2008. *Software Studies: A Lexicon*. Cambridge MA, London: The MIT Press. / Manovich, Lev. 2008. *Software Takes Command*. Accessed April 18, 2012. http://lab.softwarestudies.com/2008/11/softbook.html ↵
5. Langville, Amy N. and Carl D. Meyer. 2004. 'Deeper Inside PageRank.' *Internet Mathematics* 1:335-380. / Bianchini, Monica, Marco Gori and Franco Scarselli. 2005. 'Inside PageRank.' *ACM Transactions on Internet Technology* 5:92–128. ↵
6. Introna, Lucas D. 2007. 'Maintaining the Reversibility of Foldings: Making the Ethics (Politics) of Information Technology Visible.' *Ethics and Information Technology* 9:11–25. / Diaz, Alejandro. 2008. 'Through the Google Goggles: Sociopolitical Bias in Search Engine Design.' In *Web Search: Multidisciplinary Perspectives*, edited by Amanda Spink and Michael Zimmer, 11-34. Berlin: Springer. / Rieder, Bernhard. 2009. 'Democratizing Search? From Critique to Society-oriented Design.' In *Deep Search: The Politics of Search beyond Google* edited by Konrad

Becker and Felix Stalder, 133-151. New Jersey: Transaction Publishers. ↵

7. Zimmer, Michael. 2010. 'Web Search Studies: Multidisciplinary Perspectives on Web Search Engines.' In *International Handbook of Internet Research* edited by Jeremy Hunsinger, Matthew Allen, and Lisbeth Klastrup, 507-522. Dordrecht: Springer. ↵

8. Brin, Sergey and Lawrence Page. 1998. 'The Anatomy of a Large-Scale Hypertextual Web Search Engine.' *Computer Networks and ISDN Systems* 30:107-117. / Page, Lawrence, Sergey Brin, Rajeev Motwani, and Terry Winograd. 1999. 'The PageRank Citation Ranking: Bringing Order to the Web.' Technical Report, Stanford InfoLab. http://ilpubs.stanford.edu:8090/422/ ↵

9. Page, Lawrence. 2001. Method for Node Ranking in a Hyperlinked Database. U.S. Patent 6,285,999, filed January 9, 1998, and issued September 4, 2001 / Page, Lawrence. 2004. Method for Scoring Documents in a Linked Database. U.S. Patent 6,799,176, filed July 6, 2001, and issued September 28, 2004 ↵

10. Golumbia, David. 2009. *The Cultural Logic of Computation*. Cambridge, London: Harvard University Press. ↵

11. Galison, Peter. 1997. *Image & logic: A material culture of microphysics*. Chicago: The University of Chicago Press. ↵

12. Stigler, Steven M. 1999. *Statistics on the Table: The History of Statistical Concepts and Methods*. Cambridge, MA: Harvard University Press. ↵

13. For the wider historical and conceptual background of the methods discussed here see Mayer, Katja. 2009. 'On the Sociometry of Search Engines: A Historical Review of Methods.' In *Deep Search: The Politics of Search beyond Google* edited by Konrad Becker and Felix Stalder, 54-72. New Jersey: Transaction Publishers. ↵

14. Foucault, Michel. 1969. *L'Archéologie du savoir*. Paris: Gallimard, 83. ↵

15. Foucault 1969, 170. ↵

16. Kleinberg, Jon. 1999. 'Authoritative sources in a hyperlinked environment.' *Journal of the ACM* 46:604-632. ↵

17. Page 2004, 7. ↵

18. There are numerous PageRank implementations using different computational solutions available on the Web. Amy Langville hosts a very useful page with links to datasets and example code in Matlab: http://langvillea.people.cofc.edu/PRDataCode/index.html ↵

19. König, Dénes. 1936. *Theorie der endlichen und unendlichen Graphen*. New York: Chelsea Publishing Company, n.p. (author's translation) ↵

20. Harary, Frank. 1969. *Graph Theory*. Reading, Menlo Park, London, Don Mills: Addison-Wesley, 6. ↵

21. Harary 1969, V. ↵

22. cit. in Erdös, Paul. 1977. 'A Note of Welcome.' *Journal of Graph Theory* 1:3, 3. ↵

23. Moreno, Jacob L. 1934. *Who Shall Survive? A New Approach to the Problem of Human Interrelation.* Washington: Nervous and Mental Disease Publishing, 432. ↵

24. Moreno 1934, 9. ↵

25. Mayer 2009, 60. ↵

26. Forsyth, Elaine, and Leo Katz. 1946. 'A Matrix Approach to the Analysis of Sociometric Data: Preliminary Report.' *Sociometry* 9:340-347, 341. ↵

27. While the terms 'network' and 'graph' are often used synonymously, it should be noted that mathematicians often make a distinction based on levels of abstraction and formalization: 'There are hierarchies of mathematical models. A network is a model for empirical flows. A graph or digraph is a model for a network – it serves as a first approximation to a network with numerical or other values on its edges (or arcs) and maybe on its vertices. For a graph or digraph captures the structure of a network while suppressing the additional numerical information (…).' (Harary, Frank. 1983. 'Conditional Connectivity'. *Networks* 13:347-357, 355) ↵

28. Festinger, Leon. 1949. 'The Analysis of Sociograms using Matrix Algebra.' *Human Relations* 2:153-158. ↵

29. Chun, Wendy Hui Kyong. 2011. *Programmed Visions: Software and Memory*. Cambridge MA, London: The MIT Press. ↵

30. Bavelas, Alex. 1948. 'A Mathematical Model for Groups Structures.' *Human Organization* 7:16-30. ↵

31. Lewin, Kurt. 1936. *Principles of Topological Psychology*. New York: McGraw-Hill. ↵

32. Festinger 1949, 157. ↵

33. Barnes, John A. and Frank Harary. (1983). 'Graph theory in network analysis.' *Social Networks* 5:235-244, 237. ↵

34. Harary, Frank, and Robert Z. Norman. 1953. *Graph Theory as a Mathematical Model in Social Science*. Ann Arbor: University of Michigan. ↵

35. Harary 1969. ↵

36. Latour, Bruno. 1987. *Science in Action: How to Follow Scientists and Engineers through Society*. Cambridge, MA: Harvard University Press. ↵

37. This tendency has been particularly strong in media studies and adjacent fields where 'real' communication and computer networks proliferate. While this would seem obvious to a mathematician, when reviewing three recent titles in media studies, Berry has to explicitly state that the 'network is not ontological it is analytical' (Berry, David M. 2008. 'The poverty of networks.' *Theory, Culture and Society* 25:364-372, 365). ↵

38. Harary, Frank, Robert Z. Norman, and Dorwin Cartwright. 1965. *Structural Models: An Introduction to the Theory of Directed Graphs.* New York, London, Sydney: John Wiley & Sons, 22. ↵

39. Castells, Manuel. 1996. *The Information Age: Economy, Society and Culture. Volume 1: The Rise of the Network Society*. Cambridge MA, Oxford: Blackwell. / Benkler, Yochai. 2006. *The Wealth of Networks: How Social Production Transforms Markets and Freedom*. New Haven CT: Yale University Press. ↵

40. Sebastian Gießmann: Ganz klein, ganz groß. Jacob Levy Moreno und die Geschicke des Netzwerkdiagramms. In: Köster, Ingo und Schuster, Kai (Hg.): *Medien in Zeit und Raum. Maßverhältnisse des Medialen.* Bielefeld: transcript, 2009, S. 267-292. describes this reeducation camp for young girls, where Moreno worked, as a 'dogville in the US of the 1930s'. ↵

41. Moreno 1934, 250-251 ↵

42. Watts, Duncan J. 2004. 'The 'New' Science of Networks.' *Annual Review of Sociology* 30:243-270. ↵

43. Festinger 1949. ↵

44. Festinger 1949, 156. ↵

45. Forsyth and Katz 1946. ↵

46. Katz, Leo. 1953. 'A New Status Index Derived from Sociometric Analysis.' *Psychometrika* 18:39-43, 39. ↵

47. Katz 1953, 39. ↵

48. Katz 1953, 39. ↵

49. Katz 1953, 42. ↵

50. Katz 1953. / Hanneman, Robert A. and Mark Riddle. 2005. *Introduction to social network methods*. Riverside: University of California, Riverside. http://faculty.ucr.edu/~hanneman/ ↵

51. Hubbell, Charles H. 1965. 'An Input-Output Approach to Clique Identification.' *Sociometry* 28:377-399. ↵

52. Bonacich, Phillip. 1972. 'Factoring and weighting approaches to status scores and clique identification.' *Journal of Mathematical Sociology* 2:113-120. ↵

53. Hanneman and Riddle 2005. ↵

54. Mayer 2009. ↵

55. Garfield, Eugene. 1955. 'Citation Indexes for Science: A New Dimension in Documentation through Association of Ideas.' *Science* 122:108-111. ↵

56. Garfield, Eugene. 1972. 'Citation Analysis as a Tool in Journal Evaluation.' Science 178: 471-479, 478. ↵

57. Gross, P. L. K., and E. M. Gross. 1927. 'College Libraries and Chemical Education.' *Science* 66:385-389. ↵

58. Garfield 1972. ↵

59. Pinski, Gabriel and Francis Narin. 1976. 'Citation influence for journal aggregates of scientific publications.' *Information Processing and Management* 12: 297–312, 298. ↵

60. Pinski and Narin 1976, 298. ↵

61. Pinski and Narin 1976, 300. ↵

62. Pinski and Narin 1976, 298 ↵

63. Hubbell 1965. ↵

64. Pinski and Narin 1976, 312. ↵

65. Moreno 1934, 6. ↵

66. Gitlin's critique of the 'dominant paradigm' in communications research can therefore be productively applied to sociometric concepts as well: Gitlin, Todd. 1978. 'Media Sociology. The dominant paradigm.' *Theory and Society* 6:205-253. ↵
67. Lewin, Kurt. 1964. *Field Theory in Social Science: Selected Theoretical Papers*. New York: Harper & Row, 21. ↵
68. Mullins, Nicholas C. 1973. *Theories and theory groups in contemporary American sociology*. New York: Harper & Row. ↵
69. Granovetter, Mark S. 1973. 'The Strength of Weak Ties.' *The American Journal of Sociology* 78:1360-1380, 1360. ↵
70. Granovetter 1973, 1361. ↵
71. Emerson, Richard M. 1976. 'Social Exchange Theory.' *Annual Review of Sociology* 2:335-362, 336. ↵
72. As Emerson argues, *social exchange theory* is not a 'theory', but 'a frame of reference within which many theories – some micro and some more macro – can speak to one another, whether in argument or in mutual support' (Emerson 1976, 336). ↵
73. 'One answer, which we call "social atomism," suggests that social order is brought about through the free negotiations of autonomous individuals seeking to advance private interests. This view of social order is expressed in the "social contract" theory associated with Hobbes, Locke, Rousseau, Spencer, and others. Contemporary variations of social atomism include types of social psychological theory, especially economic-type exchange theory (Homans, 1950, 1961) (…).' (Lewis, David, and Andrew J. Weigert. 1985. 'Social Atomism, Holism, and Trust.' *The Sociological Quarterly* 26:455-471, 455) ↵
74. Mayer 2009. ↵
75. Chun 2011. ↵
76. Berry 2008, 367. ↵
77. Berry 2008, 367. ↵
78. Garfield 1972, 478. ↵
79. Porter, Theodore. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press, 74. ↵
80. Uprichard, Emma, Roger Burrows, and David Byrne. 2008. 'SPSS as an 'inscription device': from causality to description?' *The Sociological Review* 56:606-622. ↵
81. Botafogo, Rodrigo A., Ehud Rivlin, and Ben Shneiderman. 1992. 'Structural Analysis of Hypertexts: Identifying Hierarchies and Useful Metrics.' *ACM Transactions on Information Systems* 10: 142-180. ↵
82. Harary, Frank. 1959. 'Status and Contrastatus.' *Sociometry* 22:23-43. ↵
83. Botafogo el al. 1992, 143. ↵
84. Botafogo el al. 1992, 148. ↵
85. Botafogo el al. 1992, 148. ↵
86. Foucault 1969, 166. ↵
87. Mayer 2009, 68. ↵
88. Page et al. 1998, 3. ↵
89. Bianchini et al. 2005, 94. ↵
90. Chakrabarti, Soumen. 2003. *Mining the Web: Discovering Knowledge from Hypertext Data*. San Francisco: Morgan-Kauffman, 10. ↵
91. Kleinberg 1998. ↵
92. Kleinberg 1998, 2. ↵
93. Langville, Amy N. and Carl D. Meyer. 2005. 'A Survey of Eigenvector Methods for Web Information Retrieval.' *SIAM Review* 47:135-161, 136. ↵
94. Espeland, Wendy, and Mitchell L. Stevens. 1998. 'Commensuration as a Social Process.' *Annual Review of Sociology* 24: 313-343. ↵
95. Berry, Michael W., and Murray Browne. 2005. *Understanding Search Engines: Mathematical Modeling and Text Retrieval. 2nd Edition*. Philadelphia: Society for Industrial and Applied Mathematics. ↵
96. Katz 1953. ↵
97. Bonacich, Phillip. 1987. 'Power and Centrality: A Family of Measures.' *American Journal of Sociology* 92:1170-1182. ↵
98. Brin and Page 1998, 108. ↵
99. Page 2004, 7. ↵
100. Page et al. 1998, 4. ↵
101. Page et al. 1998, 5. ↵
102. Page 2004, 7. ↵
103. Page et al. 1998, 3. ↵
104. Katz 1953, 41. ↵
105. Katz 1953, 41. ↵
106. Katz 1953, 41. ↵
107. Bianchini et al. 2005, 125. ↵
108. Bonacich 1987. ↵
109. Cook, K. S., R. M. Emerson, M. R. Gilmore, and T. Yamagishi. 1983. 'The Distribution of Power in Exchange Networks: Theory and Experimental Results.' *American Journal of Sociology* 89:275-305. ↵
110. Bonacich 1987, 1174. ↵
111. Bonacich 1987, 1181. ↵
112. Bonacich 1987, 1171. ↵
113. Whether voting is necessarily a core principle of democratic government is highly debatable however. As Bernard Manin has shown, up to the French revolution, elections are generally associated with aristocracy and it is selection by lottery that is seen as the truly democratic mode of assignation of office. (Manin, Bernard. 1997. *The Principles of Representative Government*. Cambridge: Cambridge University Press.) ↵
114. Brin and Page 1998, 110. ↵
115. Diaz 2008. / Rieder 2009. ↵
116. Roush, Wade. 2004. 'Search Beyond Google.' *Technology Review*, March. Accessed April 19, 2012. http://www.technologyreview.com/articles/04/03/roush0304.2asp ↵
117. The term is used here in a rather literal sense as favoring the maintenance of the status quo. The difficulty to apply traditional concepts from political theory becomes evident when we take into account that the 'revolutionary masses' *PageRank* is set to defend against are mostly spammers and search engine optimizers. ↵
118. Katz 1953, 39. ↵
119. Introna 2007. ↵
120. Langville and Meyer 2004 / Bianchini et al. 2005 / Bechetti, Luca and Carlos Castillo. 2006. 'The Distribution of PageRank Follows a Power-Law only for Particular Values of the Damping Factor.' In *Proceedings of the 15th international conference on World Wide Web*. New York: ACM. ↵
121. In the iterative power method for calculating *PageRank*, only an approximation to the eigenvector solution is calculated. The loop stops when the changes from one iteration to the next fall below a set level. ↵
122. Langville and Meyer 2004, 346. ↵
123. Foucault 1969. ↵
124. Chun 2011. ↵