Identifying Localized Entrepreneurial Projects Through Semantic Social Network Analysis

Maria Patrizia Vittoria[(⊠)] and Pasquale Napolitano

IRISS/CNR, Naples, Italy
p.vittoria@iriss.cnr.it

Abstract. We propose a novel set of social network analysis, firm level and communication-based algorithms for mining the web to identify emerging entrepreneurial projects. These algorithms are implemented in a hybrid theoretical framework and tested in an on-line environment. The algorithms take account of entrepreneurship as the relational capability for innovation and learning, the central role of computer-mediated communication, the construction of 'dynamic' semantic networks, and the temporal computation of network centrality measures. The temporal calculation of betweenness of concepts allows us to extract and predict long-term trends for entrepreneurial projects. We illustrate our approach by considering the nodes in a network (based on our previous empirical analysis) as localized potential entrepreneurs in the cultural and creative context, and the inherent Instagram community, and analyzing the semantic networks emerging from sharing hashtags.

Keywords: Entrepreneurial project · Social media · Semantic social network analysis

JEL Codes: R11 · R15 · R58

1 Introduction

Claiming for renewed interpretation of local resource, the present paper suggests a novel set of social network analysis, firm level and communication-based algorithms for mining the web to identify emerging entrepreneurial projects. Based on a constructivist approach in social science, which seems knowledge as socially constructed, Berger and Luckmann (1969) suggest that the cognitive processes linked to entrepreneurial activities feature prominently in theories of opportunity recognition (Kirzner 1979; Beattie 1999). From this perspective, entrepreneurship can be considered a socially constructed phenomenon, which is reflected by the emergence of opportunities. These opportunities arise as individuals make sense of information and their actions and retrospectively 'discover' and 'recognize' business ideas (Gartner et al. 2003). Thus, entrepreneurship takes place in an 'enacted' environment (Weick 1995).

Compared to several other qualitative research methodologies, discourse analysis is more strongly based on social constructivist paradigm (Phillips and Hardy 2002). As known, it requires some contextualization (Cicourel 1981; Fairclough 1995); texts are the material manifestation of discourse but discourse exists beyond the individual texts

that comprise them (Chalaby 1996; Phillips and Hardy 2002). Also, a discourse cannot be identified on the basis of a single text; rather discourse emerges from the interactions among different social groups, their 'texts', and the context in which these interactions are embedded. In the case of entrepreneurship research, the context is both proximate and distal, indicating the systemic (economic) and substantive (political and cultural) embeddedness of entrepreneurship (Johannisson et al. 2002) which is reflected in the overall institutional setting, norms, and values, and the entrepreneur's political and social environment. Linking this to notions of context as discussed in discourse analysis, the proximate context and the distal context will reflect the entrepreneur's respective micro- and macro-environments (Achtenhagen and Welter 2007).

Based on selected social theories and semantic social network analysis (as a specific type of discourse analysis), we draw on the highly interconnected world of social networking platforms (Instagram[1]) to conduct an empirical exploration of a localized entrepreneurial project. We select a localized group of nodes (entrepreneurs) and visualize their Instagram community. Network centrality measures contribute to explaining the role of specific nodes-concepts-business ideas within the discourse.

The paper is structured as follows: Sect. 2 discusses the notion of entrepreneurship, Sect. 3 defines semantic social network analysis (SNA), Sect. 4 provides a synthesis of the empirical survey, Sect. 5 gives the main evidences, and Sect. 6 concludes the paper.

2 The Localized Entrepreneur and Social Media

Contemporary regional policy is increasingly interested in encouraging latent localized innovation potential in the (intentional and/or effective) entrepreneurial projects carried out by local actors (Foray 2015). We are interested in the conceptual categories that describe the basic notion of 'entrepreneurship' in this literature.

Pragmatic application of the notion of entrepreneurship inspired by development economics considers its development as a self-discovery process[2] (Hausmann and Rodrik 2003). The entrepreneurial discovery is conceived as economic experimentation with new ideas, which emanate largely from scientific and technological inventions. This chimes with the cognitive theory of the firm and its specific focus on entrepreneurship as associated with different types of learning (Nooteboom 2009). Thus, entrepreneurship can be seen as a form (individual and/or collective) of dynamic capability. It consists of the ability to find and develop external partners, which are at a

---

[1] Instagram was launched in October 2010 as a mobile photo-sharing application. It is a social network that offers its users a way to upload photos, apply different manipulation tools ('filters') in order to transform the appearance of the images, and share them instantly with 'friends' (using Instagram's application or other social networking sites such as Facebook, Foursquare, Twitter, etc.). To illustrate the pervasiveness of Instagram, in June 2013 the application had over 130 million registered users around the world who were sharing nearly 16 billion photos (Hochman and Manovich 2013). In absolute terms, although less diffused than some other social media, Instagram is more popular with high skilled Internet users, and with women.

[2] It is a particular type of learning: learning 'what one is good at producing'.

distance[3], and the intellectual and behavioral capability to collaborate across this distance. It takes account of both competence and governance issues (Nooteboom 2009).

It is a fact that the new opportunities from participation in open-source communities and social networking platforms are contributing to a more complex notion of the entrepreneur. We are interested in the effects of textual meta-data matrices within the communication on social networks, in particular Instagram, and how these linguistic signs - verbal language fragments in a social platform strictly visual based - can express meaning for research especially from the perspective of the relational dimension of the focal network.

The approach to language introduced by Austin (1962) with his definition of performative utterance, suggests that language should no longer be considered a descriptive tool related to a state of affairs but should be understood as an act of creation - "performative" - of the real. Evoking the categories of thought through language is to create meaning, in the case of hashtags, commentary and description textual meta-data of the visual products issued by the users of the social. For each user it means building an individual identity, an individual biography. These individual biographies mediated by social-media when if shared, evolve into a value that is a more complex system which transcends the individual dimension of the individual user and is shared by multiple users[4].

The novelty here is that, in the utterance, the individual is performing an action of which the very act of uttering the sentence is an essential component. We propose to introduce *performative utterance* in an assessment of a corporate network empirically. This allows us to analyze what is imprinted in the statements of identity discourse generated through the hashtag, on the Instagram profiles of community actors, and what it means when placed in a relational intentionality typical of network dynamics. Our interest is in identifying the self-representation produced by the meta-data, and to investigate which metadata are most commonly shared by the actors in the network. The analysis is conducted in two phases to examine the universes of identity, values, and interests that characterize the network and its actors. This research extends the notion of entrepreneurship as the ability to find and develop outside partners at sufficient cognitive distance. Additional relational competencies are needed, and particularly the ability to compete in global knowledge networks.

---

[3] The way this literature considers the concept of 'distance' is interesting. It is not seen as fixed distance between two situated nodes (geo-graphical distance) but as a 'cognitive' gap, or a variable space resulting from the trade-off between the novelty value of cognitive variety and the need for some cognitive coherence to utilize the potential for novelty.

[4] These are precisely the forms of transmission, including the writing of social media, which determine the ability to generate ideas, because in this context, every statement is emancipated from its subjective and contingent character, to become a form of collective writing (Husserl 2015). The storytelling that daily produces a user through social is configured as a performative narrative, based on operational narratives of more performances.

3 Semantic Social Network Analysis

Grounded in the field of communication science, semantic network analysis (Popping 2000) can be considered an alternative to content analysis (CA) (Krippendorff 2004). Since CA is used to analyze the content of media messages, it tends to determine the value of one or more variables based on the message content. In other words, it infers relevant aspects of what a message (newspaper article, forum posting, personal e-mail, etc.) means in its context, and the communication research question determines both the relevance and the correct context.

Rather than directly coding the messages to address the research question, semantic network analysis first represents the content of the messages as a network of objects. This network representation is queried to address the research question.

Despite wide use of the technique, extracting the network of relations from the text can be more difficult than categorizing text fragments although there are no standards for defining patterns on these networks (van Atteveldt 2008).

New social media such as Facebook, Twitter, Instagram, and so on, are considered direct and indirect big relational data sets. Within these virtual places, huge amounts of content (photo and/or video posts and blogs) are shared socially at diverse levels with different motivations such as socializing, co-designing, etc. This kind of social sharing is considered semantic due to the nature of the shared objects. The strength of this kind of on-line semantic sharing lies in the network structure and in its power of viral transmission of the messages/content[5].

Even more trans-disciplinary technique, actually, there are three implication levels, as scientific fields directly involved in.

Firstly, 'computational linguistics' has seen drastic increases in computer storage and processing power in recent decades, leading to the development of multiple linguistic tools and techniques. Second, there is a need to alleviate the problems of combining, sharing, and querying these semantic networks, which requires a focus on 'knowledge representation'. This refers to the formal representation of the background knowledge used to aggregate the textual objects with the abstract concepts in a research question. Third, there can be the distinguishing manual and automatic extractions of complex and abstract concepts by these data sets[6]. A frequent application of automatic extractions is marketing trend analysis and political science. At this level, the basic research question is about measuring the concept's relative importance in the relevant information sphere (web, blog, on-line forum). If the concept (e.g. a hashtag) is a node in a network of links (e.g. sharing hashtag), then analysis of the network structure can reveal the relative importance of that concept. Thus, semantic SNA is an extension of the SNA method (Wasserman and Faust 1994). The concepts of high betweenness

---

[5] In these respects, Barabasi (2003) was pioneering research.

[6] Condor is a sophisticated semantic SNA tool (Gloor et al. 2009). It includes automated textual analysis functionality using standard information retrieval algorithms such as 'term frequency-inverse document frequency'. Also, it factors in the betweenness centrality of nodes to weigh the content by the social network position of the nodes.

centrality (BC) (the semantic, more diffused SNA indicator) become gatekeepers between different domains[7].

4   Exploring an Entrepreneurial Project (Ep) Through Semantic SNA: A Synthesis of the Empirical Survey

The empirical analysis is in two steps. The first is an interpretative firm-level case study to identify a localized Ep project[8].

The evidence includes the multi-relational external networks, corresponding to a specific learning investment.

Starting from these networks (A ≡ 6 nodes and 8 links; B ≡ 5 nodes and 8 links), we can parse the corresponding Instagram communities (research step 2).

The research dataset consists of the hashtags[9] emitted in the previous two years by all members (6 + 5) of the networks related to the case study.

The research on the Instagram database involved several stages: search of the content was enabled using the tool Iconosquare (http://iconosquare.com/) which, only giving information on the user's Instagram account, provides more objective research content since it is free of local and temporal constraints, which constrain search performed by users directly approaching a company.

In a subsequent step, data collection consists of gathering company information on companies and compiling it in a database record using the "trans-coding" language (Manovich 2001) 'python' which is a script that can extract data from Instagram through the API protocol[10].

---

[7] SNA provides a lot of measures for quantifying a member's interconnectedness within social networks. As is well known, each indicator can be critically analyzed according to its explanatory capabilities and the context of analysis (cfr. Landherr et al. 2010). If we consider centrality as the control of the information flow that a member of a network may exert based on his position in the network, the concept of BC is on the track. It is given by the quotient of the number of all shortest paths between actors in the network that include the regarded actor and the number of all shortest paths in the network (Freeman 1977).

[8] The case observed is the Fondazione Plart located in Naples, Italy. It is dedicated to recovery, restoration, and conservation of artifacts and design objects constructed from synthetic materials (plastics). It is a research and restoration lab, an event location, a training center, and a permanent exhibition site for its founder's historical plastics collection. The Plart's founder is largely recognized as a potential entrepreneur in the local cultural and creative context.

[9] It was decided to analyze the hashtags emitted in a 2-year period in order to even out the differences related to the longevity of the different Instagram profiles as well as the various geographical locations. We chose a period when all profiles and geographical locations were on the company in order to allow comparative analysis which would be chronologically balanced.

[10] Instagram provides API, i.e. the protocols for querying to extract information about the data, to be used in external applications such as external analytics tool platforms. The link below identifies all the endpoints, i.e. all types of calls carried including a direct link to the images, its statistics (likes and comments), and the set of accompanying text which then provides all the hashtag used - https://www.instagram.com/developer/endpoints/.

Before moving on to the phase of data mining, we performed some cognitive ergonomics operations aimed at avoiding redundant or insignificant data as follows:

1. We deleted from the dataset the "auto-tag", i.e. all those hashtags in which the issuer of the media content "tags you". This initial screening is necessary to avoid imbalanced data with respect to individuals who issue more content, and especially because the self-tag may not be construed as a given relational potential, given its self-referentiality;

2. We deleted from the dataset all the "omnibus" hashtags, i.e. all those metadata which act as a description or interpretation of the photographic content that accompanies them, performing a "channel function", i.e. referring to Instagram and its practices and operations. This category includes the meta-data "#instagram" "#in-stantmood" "#instantgood" "#igers" "#photosofthedayh" "#pi-coftheday" "#vso" "#vsocam" "#tagsforlikes" etc. This second filter was necessary to avoid drugging the results of the sampling with metadata not related to the specific of the analyzed subject but present in all or most of the Instagram content, a hashtag shared around the world used by users as a tool to cope with a greater amount of feedback and a greater rate of engagement. The dataset was split into two, coinciding with the hashtag inherent in the actor-network active first in the exploration phase (network A, with 6 detected users) and then in the exploitation phase (network B, with 5 detected users), within the learning cycle of the case study. To avoid redundancy, we aggregated the meta-data conceptually. To reduce redundancy and increase truth in substantive terms, we proceeded to a tag aggregation, both lexical proximity ("Naples" and "Napoli", "Milan" and "Milano", "graphic" and "grafica") and conceptual proximity (combinations of words that are synonymous or referring to a genus of the same species, e.g. "Arduino" and "Arduinolab"). We obtained a dataset of 734 records - 274 for the exploration phase, and 470 for the exploitation phase. Data visualization was achieved through a dual network visualization where the shared hashtags serve as bonds (marked with a circle) between network nodes (the actors in the network, identified by a square) (see Fig. 1).

Finally, to the aim of implementing the data analysis the research of proper centrality measures (i.e. Betweenness Centrality of vertices in complete Bipartite Graphs) could be absolved (cfr. Unnithan et al. 2014).
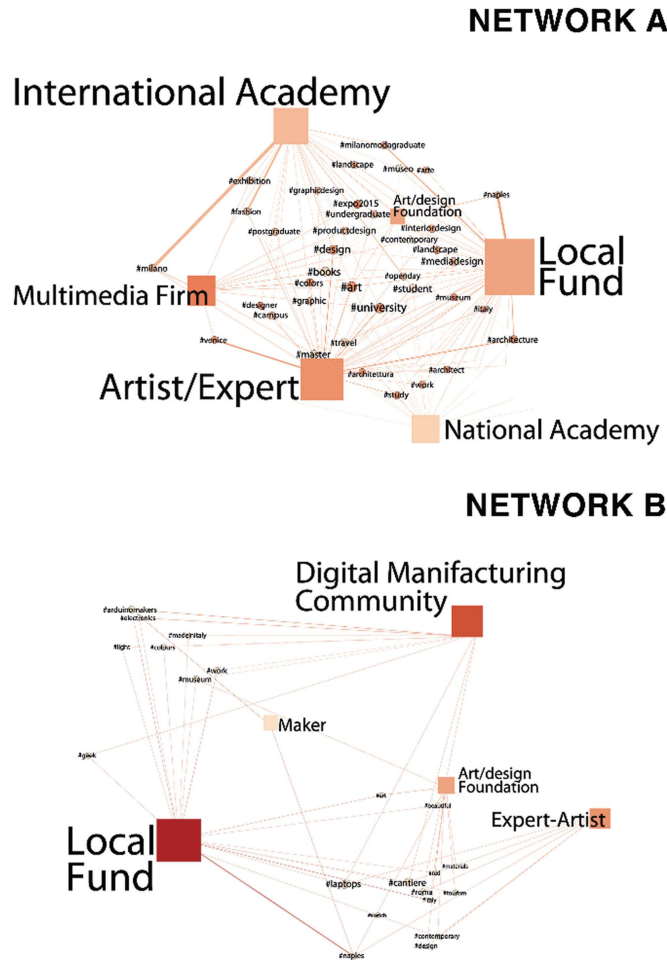
NETWORK A

International Academy

Multimedia Firm

Art/design Foundation

Local Fund

Artist/Expert

National Academy

NETWORK B

Digital Manifacturing Community

Maker

Art/design Foundation

Expert-Artist

Local Fund

Fig. 1. Comparison - between Phase 1 (above) and Phase 2 (below) - of the importance in Instagram platform of the "shared hashtags" for each phase. Squares are Instagram users, circles are hashtags, and node size denotes betweenness.

5   Main Evidences

Results from the first research step (case study) are given by the description of the Fondazione Plart Ep. It is related to its dynamic abilities to capture the largest number of experts in the field of polymeric materials and the strategies of external consulting agreements, and indirect involvement of artists and experts in exhibition activities. The involvement of this expertise (and thus, structured inter-organizational links) at different degrees of cognitive distance, allows a larger training supply (e.g. in-depth teaching and laboratory activities replicable at home and in school for younger people).

Results from the second research step emerge by the semantic network analysis of a selected group of Instagram users formed by Fondazione Plart's external network nodes. A main evidences' synthetic view is in Fig. 1, where the network visualizations are scaled according to the betweenness centrality values[11]. This provides an understanding of which occurrences are the most shared by the users of the network in each phase.

When interpreting the data, it was clear that the evolution of the cycle of learning, the network concerned, at least as regards its narration on social media, evolves into an equally sharp. In the first phase, the network aggregates of semantic fields are related to architecture and design disciplines - all members of the network share the #architecture and #design hashtag. In the second phase, the network also aggregates on a geographical basis, and extends its domain epistemic to the world and the art market, tourism, new creative forms, such as the universe of 'making'.

6   Conclusion

This research has demonstrated that selecting a group of nodes that are connected by a learning logic based on cognitive distance allows social media (Instagram) to be used as a source of information for research in entrepreneurship.

This study contributes to our understanding of complex social networks by studying the modular structures of networks. Detecting the network modules, or communities, is becoming a critical issue and there is much discussion on the quality of the partition process. Our research testifies to the importance of aligning the research question and its theoretical background (finding a localized entrepreneurship project as a self-discovery process), with the research method (explorative), and thus, fixing the algorithm for mining the web (selecting Instagram users through a firm-level case study). To model the formation of a community (in our conceptual background) we used a learning-based entrepreneurial process, which treats each node as a player in a heuristic of invention (a dynamic cycle alternating the phases of exploration and exploitation).

---

[11] In particular, it is possible to scale the values - for each phase - according to two statistical categories. I.e. we can take account of the phenomena of node degree (which correspond to the 'what' kind of information) and betweenness centrality (which approximates the influence of a concept, the 'who' kind of information, within a discussion) for each node and each link.

References

Achtenhagen, L., Welter, F.: Media discourse in entrepreneurship research. In: Neergaard, H., Ulhøi, J.P. (eds.) Handbook of Qualitative Research Methods in Entrepreneurship, pp. 193–215. Edward Elgar, Cheltenham, Northampton (2007)

Austin, J.L.: How to Do Things with Words. Oxford University Press, Oxford (1962)

Barabási, A.-L.: Linked: Linked: how everything is connected to everything else and what it means for business. Science and Everyday Life. Plume 2 (2003)

Beattie, R.: The creative entrepreneur: a study of the entrepreneur's creative processes (1999). www.babson.edu/entrep/fer/papers99/III/III_B

Berger, P.L., Luckmann, T.: Die gesellshaftliche Konstruktion der Wirklichkeit: Eine Theorie der Wissenssoziologie. Fischer, Frankfurt (1969)

Chalaby, J.K.: Beyond the prison-house of language: discourse as a sociological concept. Br. J. Sociol. **47**, 684–698 (1996)

Cicourel, A.V.: Three models of discourse analysis: the role of social structure. Discourse Process. **3**, 101–131 (1981)

Fairclough, N.: Media Discourse. Edward Arnold, London (1995)

Foray, D.: Smart Specialization. Opportunities and Challenges for Regional Innovation Policy. Routledge, Taylor & Francis Group, London, New York (2015)

Freeman, L.C.: A set of measures of centrality based on betweenness. Sociometry **40**(1), 35–41 (1977)

Gartner, W., Carter, N., Hill, G.: The language of opportunity. In: Steyaert, C., Hjorth, D. (eds.) New Movements in Entrepreneurship, pp. 103–124. Edward Elgar, Cheltenham, Northampton (2003)

Gloor, P.A., Krauss, J., Nann, S., Fischbach, K., Schoder, D.: Web Science 2.0: identifying trends through semantic social network analysis. IEEE Computer Society (2009)

Hausmann, R., Rodrik, D.: Economic development as self-discovery. J. Dev. Econ. **72**, 603–633 (2003)

Hochman, N., Manovich, L.: Zooming into an Instagram City: reading the local through social media. First Monday **18**(7) (2013). https://doi.org/10.5210/fm.v18i7.4711

Husserl, E.: La crisi delle scienze europee e la fenomenologia trascendentale. Saggiatore, Milano (2015). traduzione di E. Filippini, prefazione di E. Paci, il [1961]

Johannisson, B., Ramirez-Pasillas, M., Karlsson, G.: The embeddedness of inter-firm networks. Entrepreneurship Reg. Dev. **4**(14), 297–315 (2002)

Kirzner, I.: Perception, Opportunity and Profit: Studies in the Theory of Entrepreneurship. University of Chicago Press, Chicago (1979)

Krippendorff, K.: Reliability in content analysis: some common misconceptions and recommendations. Hum. Commun. Res. **30**(3), 411–433 (2004)

Landherr, A., Friedl, B., Heidemann, J.: A critical review of centrality measures in social networks. Bus. Inf. Syst. Eng. **2**(6), 371–385 (2010)

Manovich, L.: The Language of New Media. The MIT Press, Cambridge, London (2001)

Nooteboom, B.: A Cognitive Theory of the Firm. Learning, Governance and Dynamic Capabilities. Edward Elgar, Cheltenham, Northampton (2009)

Phillips, N., Hardy, C.: Discourse Analysis: Investigating Processes of Social Construction. Sage, London, Thousand Oaks (2002)

Popping, R.: Computer-Assisted Text Analysis. Sage, London (2000)

Unnithan, S., Kannan, B., Jathavedan, M.: Betweenness centrality in some classes of Graphs. Int. J. Comb. **2014**, 1–12 (2014). ID 241723

Van Atteveldt, W.: Semantic network analysis. Techniques for extracting, representing, and querying media content. Vrije Universiteit (2008). http://vanatteveldt.com/dissertation

Wasserman, S., Faust, K.: Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge (1994)

Weick, K.E.: Sensemaking in Organizations. Sage, London, Thousand Oaks (1995)