LLNL-CONF-681917

LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

# Characterizing Microseismicity at the Newberry Volcano Geothermal Site using PageRank

A. C. Aguiar, S. C. Myers

February 4, 2016

Stanford Geothermal Workshop
Stanford, CA, United States
February 22, 2016 through February 24, 2016

PROCEEDINGS, Fourtieth Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 22-24, 2016
SGP-TR-209

# Characterizing Microseismicity at the Newberry Volcano Geothermal Site using PageRank

Ana C. Aguiar, and Stephen C. Myers

Lawrence Livermore National Laboratory, 7000 East Ave L103, Livermore, CA 94550-9234

aguiarmoya1@llnl.gov, and myers30@llnl.gov

**Keywords:** Microseismicity, EGS, Newberry, PageRank

## ABSTRACT

The Newberry Volcano, within the Deschutes National Forest in Oregon, has been designated as a candidate site for the Department of Energy's Frontier Observatory for Research in Geothermal Energy (FORGE) program. This site was stimulated using high-pressure fluid injection during the fall of 2012, which generated several hundred microseismic events. Exploring the spatial and temporal development of microseismicity is key to understanding how subsurface stimulation modifies stress, fractures rock, and increases permeability. We analyze Newberry seismicity using both surface and borehole seismometers from the AltaRock and LLNL networks. For our analysis we adapt PageRank, Google's initial search algorithm, to evaluate microseismicity during the 2012 stimulation. PageRank is a measure of connectivity between an instance (web-site, event, signal, etc.) and a collection of other instances, where higher ranking represents more connections. In our seismic application connectivity is measured by the cross correlation of 2 time windows recorded on a common seismic station and channel. Aguiar and Beroza (2014) used PageRank to detect low-frequency earthquakes, which are highly repetitive but difficult to detect. We expand on this application by using it to define signal-correlation topology for micro-earthquakes, including the identification of signals that are connected to the largest number of other signals. We then use this information to create signal families and compare PageRank families to the spatial and temporal proximity of associated earthquakes. Studying signal PageRank will potentially allow us to efficiently group earthquakes with similar physical characteristics, such as focal mechanisms and stress drop. Our ultimate goal is to determine whether changes in the state of stress and/or changes in the generation of subsurface fracture networks can be detected using PageRank topology.

## 1. INTRODUCTION

Enhanced geothermal systems (EGS) have the potential to be an important source of renewable energy in the future. Understanding how these systems work is crucial to the development of this technology. The Newberry Volcano located south of Bend in Central Oregon, was selected by AltaRock Energy and Davenport Newberry to test and help demonstrate the EGS technology (Cladouhos *et al*. 2012). After an extensive study of the state of stress for the area (Cladouhos *et al*. 2011a; Davatzes and Hickman 2011), this location was deemed to have a very low permeability rate as a well as a large conductive thermal anomaly that yields high-temperatures (Cladouhos *et al*. 2011b), making it ideal to test as an EGS.

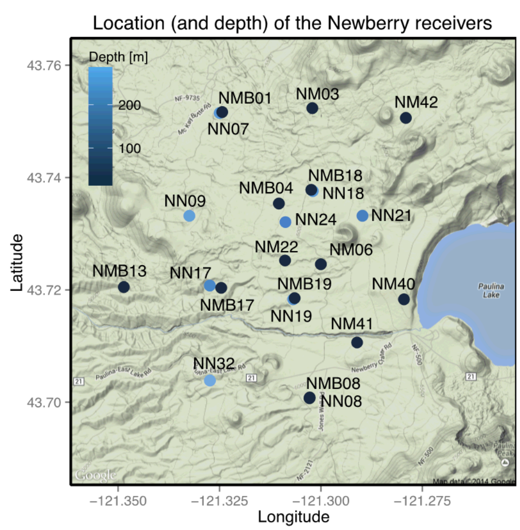


**Figure 1: Location Map of the Newberry EGS site and stations used in the analysis. Color indicates the depth at which the receivers are installed where dark blues are shallow stations and light blues are deeper stations.**

Part of the project proposed by AltaRock consisted in stimulating well NWG 55-29 (close to station NM22 on the west flank of the Volcano), between October 2012 and February 2013, to induce hydroshearing (Cladouhos *et al*. 2011b; Cladouhos *et al*. 2012). The stimulation induced a series of microseismic events, with the first event detected October 29, 2012. The seismicity continued for two months until the last event in the area was reported on January 2, 2013 (Cladouhos *et al*. 2013). From this stimulation, AltaRock created a catalog of 226 microseismic events.

To complement the existing Pacific Northwest Seismic Network stations around the volcano, AltaRock (as well as the Lawrence Livermore National Laboratory) installed a large number of seismic receivers around the area of interest (Figure 1), containing both surface and borehole seismometers at different depths. Given the good data coverage of the study site, the Newberry volcano is an ideal place to test new methods for characterizing microseismicity. In this study we want to analyze the microseismicity by applying a data mining method to learn more about the events caused by the stimulation and explore the extent to which we can describe the physical characteristics of the microseismic events. We apply Google's original search algorithm, PageRank (Page *et al*. 1999), which was initially developed to determine the ranking or "importance" of a webpage based on the number of links it has to other highly ranked pages. In EGS sites, microseismicity is assumed to illuminate injection flow paths while hydroshearing occurs, but it can also be related to changes in fluid pressures in other areas away from the injection well. By applying PageRank to this microseismicity data set, we can analyze how events are linked in space and time, helping us detect changes in microseismic signals that may indicate changes in the state of stress as well as changes in pore pressure in the areas surrounding the injection well.

## 2. THE PAGERANK APPROACH

Waveform correlation coefficient (CC) is a common, robust metric used in seismology for linking previously recorded signals and their associated events. Although, in an autocorrelation scheme, where waveforms may not be available and each section of the data could be used as a potential template, positive matches do not necessarily satisfy "closure." That is, if event A correlates with event B later in time, and event B correlates with another event C also later in time, then event A and event C do not necessarily correlate with one another (Figure 2). This problem is directly addressed by Google's initial search algorithm (Page *et al*. 1999) called PageRank.

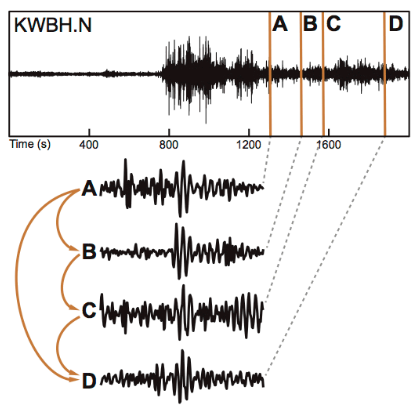


**Figure 2: From Aguiar and Beroza (2014). Window pair links: A, B, C, and D are windows that were found to be a match with another window, given the thresh- old selected. This figure shows how all these pairs are linked together. Some windows are linked directly: A → B and B → C. Other windows are linked indirectly: A is linked indirectly to C (be- cause B → C), and B is indirectly linked to D (because C → D).**

Aguiar and Beroza (2014) originally applied PageRank in seismology to exploit the repetitive nature of low frequency earthquakes (LFEs) within continuous seismic data. For the case of the LFEs, PageRank facilitated the analysis of all events linked, both directly (events A and B in Figure 2) and indirectly (events A and C in Figure 2), to a reference event with high PageRank. Including the indirectly linked events in the analysis made it possible to create more robust templates and use them in continuous detecting schemes, by greatly improving the signal to noise ratio of the template (Aguiar and Beroza, 2014).

PageRank, as developed by Page *et al*. (1999) for webpages, is the probability that a "random surfer" will visit a particular web page. Originally, PageRank would assign a ranking for a webpage based on the number of links associated with that page, where, higher ranking represents pages with large numbers of links to other pages with many links. PageRank is an iterative process represented by the following equation:

$$\vec{x} = \mathbf{A}\vec{x} \tag{1}$$

Here, the elements of the transition probability matrix **A** are considered as follows

$$a_{ij} = \begin{cases} pg_{ij}/c_j + \delta & \Rightarrow\ c_j \neq 0 \\ 1/n & \Rightarrow\ c_j = 0 \end{cases} \tag{2}$$

and this matrix is where all the link information between events is contained. Here $c_j = \sum_i g_{ij}$ and $\delta = (1-p)/n$; and in equation (1) $\vec{x}$ represents the PageRank, $g_{ij}$ are the links between windows (where $g = 1$ if there is a link and $g = 0$ otherwise), $n$ is the total number of windows, $p$ is the probability of following a link and $\delta$ is the probability of finding yourself on a certain random window. The transition matrix is a very sparse matrix, therefore making it a non- intensive computational process.

### 2.1 PageRank Applied to Newberry EGS Microseismicity

In the case of microseismicity from a geothermal site, given a common source, many of the events will be close in time and space and potentially show similar physical characteristics. By applying PageRank, we can analyze the connectivity of this data set, not just for the direct links between events, but also for the indirect links. We apply PageRank to all data from both surface and borehole seismometers (AltaRock and LLNL) for the 226 events in the data set, and find all links possible for each highly ranked event. We focus here on two of the best linked events, shown in Figure 3, that present two very distinct clusters in the data.

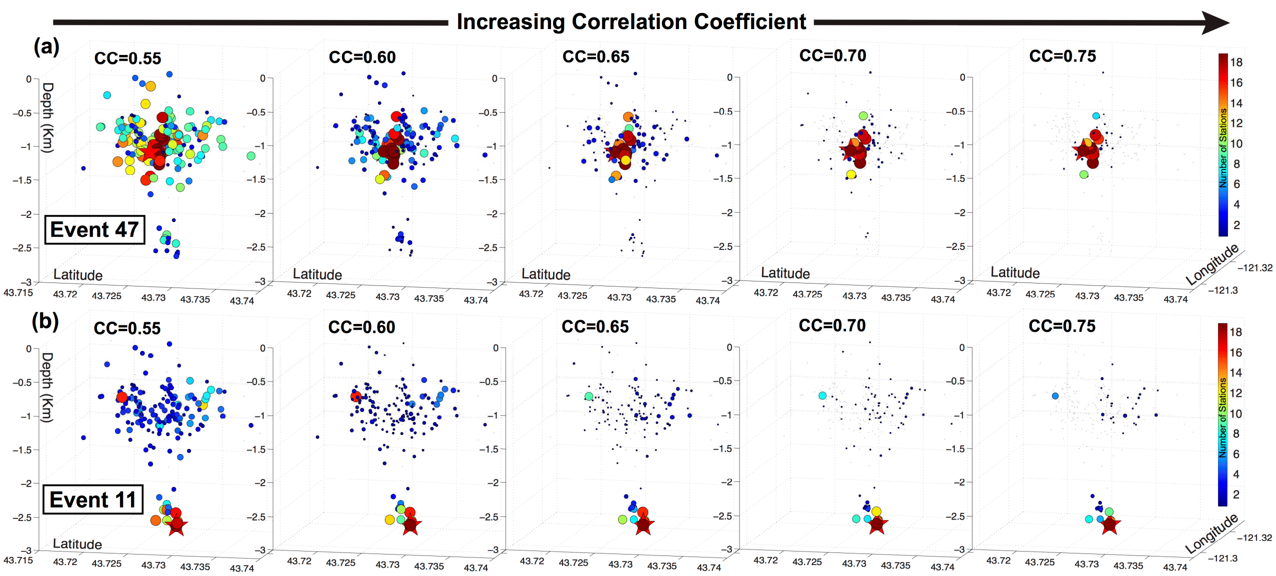


**Figure 3: Result for increasing detection threshold of events linked to two of the high PageRank events found in the analysis. (a) Plots for reference event 47 of the analysis (Red Star). (b) Plots for reference event 11 of the analysis (Red Star). Color and size of each event indicates the number of stations that link that event to the reference event. That is, larger and warmer color events were linked to the reference event by a larger number of stations.**

For a very low detection threshold (CC=0.55), most of the events are link at at least one station (Figure 3, initial panels for a and b). As we increase the CC threshold in the analysis we see that many of the events start to lose their link to the reference event (Figure 3). But, if the threshold is too high, linkage information can be lost. Traditional cross-correlation can miss a large number of detections due to the use of higher thresholds, which are represented as the small gray stars in both sequences in Figure 3. However, by requiring linkage at a large number of stations (as in this example: 6 or more, the warmer colors in the plot), PageRank allows us to lower the CC threshold significantly, and still detect many events that would have been previously missed with a high CC threshold. This also indicates that it is necessary to observe an event in a large number of stations to confidently determine that the events – not just signals – are linked.

## 3. DIRECTLY AND INDIRECTLY LINKED EVENTS

As previously mentioned, PageRank allows us to determine indirectly linked signals. By doing so, we can lower our CC threshold and maximize the number of events found that are determined to be related to one another. To show the advantage of including indirectly linked windows in the analysis we compare these results to the analysis using only direct links. Figure 4 shows this test for the events linked to reference event 47.

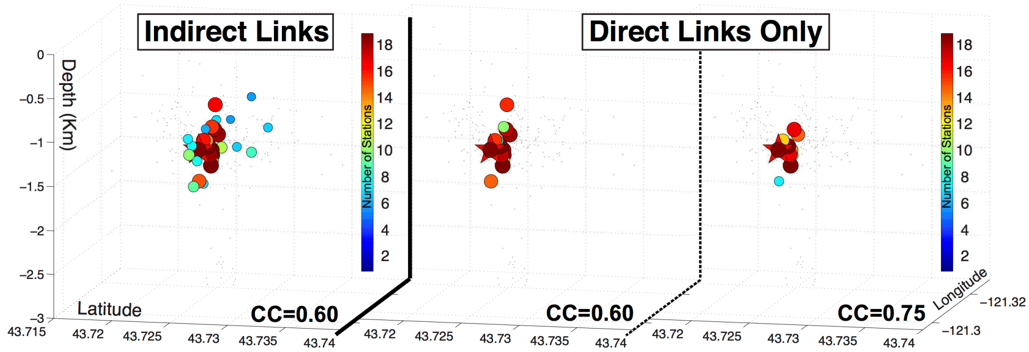


**Figure 4: Comparison between results including both direct and indirect linked events (left) for a threshold of CC=0.6 and including only directly linked events (middle and right) for detection thresholds of CC=0.6 and CC=0.75. Again, color and size represent the number of stations that link the event to reference event 47. Warmer colors and larger size both represent a larger number of stations as indicated in the color bar. The red star represents the reference event (high PageRank).**

We also find that to efficiently and confidently do this, we need to require signal linkage at a large number of stations to declare that the events themselves are linked. That is, confidently linking events requires linkage of signals at a large number of stations at a wide range of azimuths. Figure 4 shows how many indirectly linked events are also clustered around the reference event (red star) as linked at a large number of stations (in this example, 6 or more stations are required for detection).

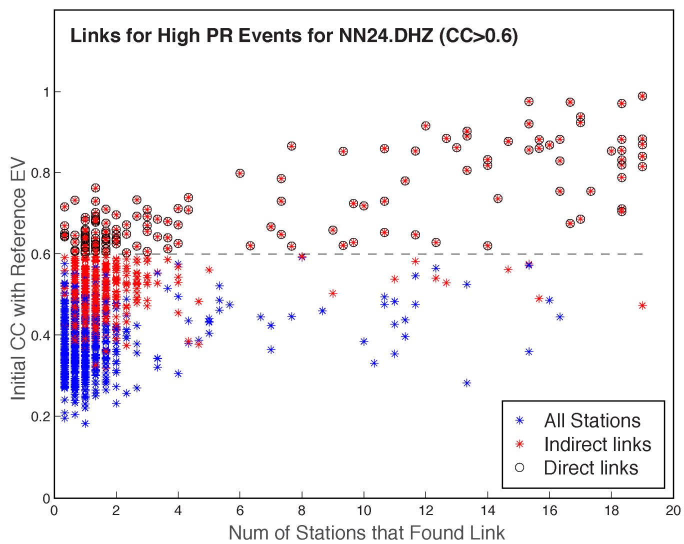


**Figure 5: CC at station NN24 vs. number of stations linking the event to the reference event. Direct links (red stars with black circles), indirect lines at station NN24 (red star), indirect link at other stations (blue stars).**

We can also show that, for a specific threshold (CC=0.6 in this case), the number of indirectly linked events (below the detection threshold) found by a higher number of stations is very large. This is shown in Figure 5 for station NN24. We see that for CC< 0.6 (below the dashed line in Figure 5) there are many indirectly linked events at stations other than NN24. As expected, many of these links are made at a limited number of stations (the large cloud with links at 1 to 3 stations in Figure 5). But our confidence that the events are linked increases as events are linked at more stations, even though the signals at station NN24 were not highly correlated. In some cases, direct or indirect links are made at as many as 19 other stations (blue stars) even though the CC at station NN24 is below the 0.6 threshold. This shows that these indirectly linked events, even though missed by a standard cross-correlation detection, are still meaningful detections. Because of this, we concluded that the most sensitive and robust metric for linking events at the Newberry site is to include PageRank in the analysis, allowing us to determine indirect links and require direct or indirect links at a at least 6 stations to declare a true linked event.

### 3.1 Microseismicity Waveforms

Analyzing the data in detail, we can show how the indirectly linked event waveforms are good matches to the reference event, even though their cross correlation to the reference event did not exceed the CC detection threshold. Figure 6 shows an example for 3 stations used in the analysis with several events (black) indirectly linked to reference event 47 (red). All of the indirectly linked events plotted in Figure 6 were found in at least 6 of the stations used in the analysis.

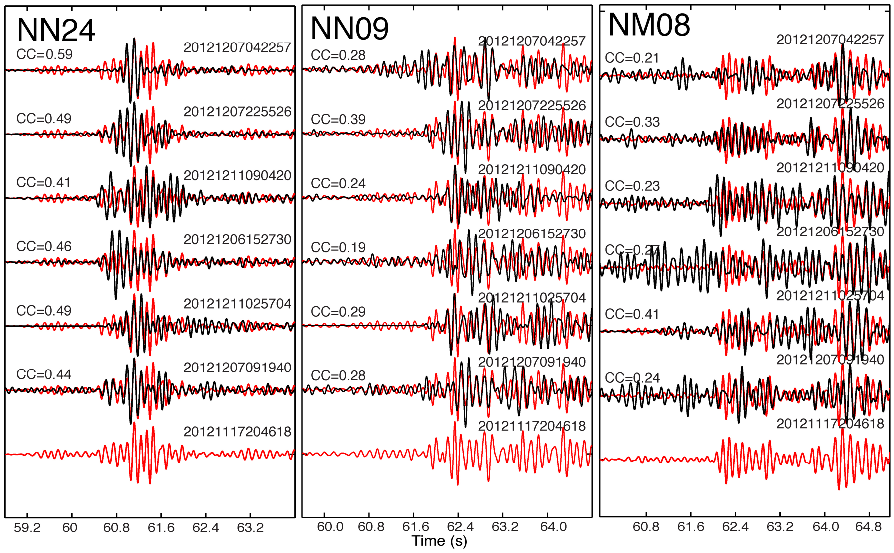


**Figure 6: Three different stations showing the seismic signal of six different events found indirectly linked (black traces) to reference event 47 (red traces).**

Figure 7 on the other hand shows the waveforms for one event (event 141) in black whose CC at station NN17 was 0.69, thus exceeding the CC threshold (direct link) (fourth trace from the top).

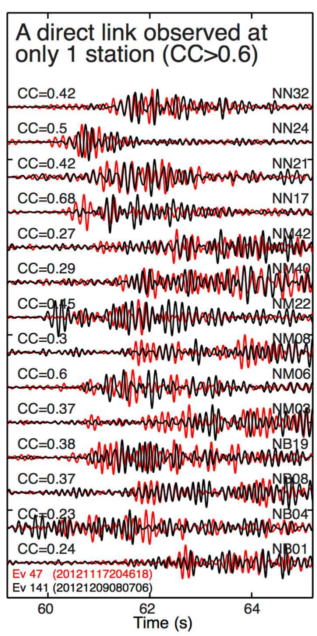


**Figure 7: Plot for all stations showing seismic signal for event 141 (black traces) found as a direct link (CC above initial detection threshold), in <u>only</u> one station, to reference event 47 (red traces).**

Even though a direct link is established at station NN17, signals at most other stations are neither directly nor indirectly linked. This shows the importance of network-wide assessment of event linkage and the utility of using indirect links.

## 2. CONCLUSIONS

Traditionally, waveform correlation coefficient (CC) is a common metric for linking signals and their associated events, which is what we have designated as direct link waveforms. Linking events is commonly used to infer that events are close to one another and have similar physical properties, such as focal mechanism.

In this study, PageRank extends signal linkage to include the similarity of two signals with a third signal, or what we have designated as an indirect link. We have shown that events are in close proximity to one another only when their signals are linked directly or indirectly at many stations. Including the indirectly linked events to the analysis gives us a more robust result as more close proximity events are linked when indirect links are included in the analysis.

Focusing on this extended linkage between events will allow us to more efficiently find all events that show very similar physical characteristics, and in turn let us analyze the microseismic process, in space and time.

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

## REFERENCES

Aguiar, A.C. and Beroza, G.C.: PageRank for Earthquakes, Seismological Research Letters, **85**, Number 2, (2014), 344-350.

Cladouhos, T.T., Clyne, M, Nichols, M., Petty, S., Osborn, W., and Nofziger, L.: Newberry Volcano EGS Demostration Stimulation Modeling, GRC Transactions, **35** (2011a), 317-322.

Cladouhos, T.T., Petty, S., Callahan, O., Osborn, W., Hickman, S. and Davatzes, N.: The Role of Stress Modeling in Stimulation Plannig at the Newberry Volcano EGS Demostration Project, *Proceedings*, 36th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2011b).

Cladouhos, T.T., Osborn, W.L., Petty, S., Bour, D., Iovenitti, J., Callahan, O., Nordin, Y., Perry, D., and Stern, P.L.: Newberry Volcano EGS Demonstration—Phase I Results, *Proceedings*, 37th Workshop on Geothermal Reservoir Engineering, Stanford University, Stanford, CA (2012).

Cladouhos, T.T., Petty, S., Nordin, Y., Moore, M., Grasso, K., Uddenberg, M., Swyer, M., Julian, B., and Foulger, G.: Microseismic Monitoring of Newberry Volcano EGS Demonstration, *Proceedings*, 38th Workshop on Geothermal Reservoir Engineering Stanford University, Stanford, CA (2013).

Davatzes, N., and Hickman, S.H.: Preliminary Analysis of Stress in the Newberry EGS Well NWG 55-29, GRC *Transactions*, **35** (2011), 323-332.

Page, L., Brin, S., Motwani, R., and Winograd, T.: The PageRank citation ranking: Bringing order to the Web, *Technical Report*, Stanford InfoLab (1999).