Implementation Plan (Forks vs Citation)

**Goal**: To implement the Disruption formula *given below* with Linux repository as the reference      

*ni*– *nj*

Disruption  =      ---------------

*ni*+ *nj*+ *nk*

*ni  = Number of subsequent works citing the focal paper*

*nj  = Number of subsequent works citing both focal paper and references*

*nk = Number of subsequent works citing references*

In Linux Repository terms:

ni = Number of subsequent works citing Linux repository.

nj = Number of subsequent works citing both Linux repository and references of a Linux repository

nk= Number of subsequent works citing references.

To implement the formula we need to recognize the subsequent works, focal paper and references

**Analogy:**

**Focal Paper:** Linux Repository

**References: ?**

**Subsequent work: ?**

**We need to identify the references and subsequent work of Linux repository.**

**How?**

To identify the references, subsequent work and the citations between them we need to build a **citation network with reference to Linux repository**.

**Rule to draw a citation link between repositories:**

**We can draw a citation link between repository A and B if the core contributor of A has forked B while he was actively working on A.**

**Understanding terminology:**

**Core Contributor of a repository:** If the number of edits or pushes of the contributor is greater than the average number of edits or pushes of all the contributors to a project. He is said to be a Core Contributor of that particular repository.

**Active period on a repository:** The time period between the first and last edit of the core contributor on a repository.

**Sample citation network for Linux repository:**

Let us assume that **CC** is the core contributor for Linux repository. We divide our structure into four parts:

**Part A**

* U, V, X, Y, Z are the repositories CC has forked.
* But CC has forked only X, Y, Z while he was working on Linux repository.
* Now that our condition to draw a citation link is satisfied, we can draw citation links from **X, Y, Z to Linux**.
* We can infer that X, Y, Z are the references to the Linux repository and that Linux cited X, Y, Z.

X

Linux

Z

Y

**Fig:1 Part A**

**The above figure is the citation network formed for the given repositories**

**Part B**

Let us assume that the contributors of Linux are given by Cl.

* The contributors of Linux be Cl1, Cl2 and they have {R1, R2, R3}, {R4, R5, R6} as corresponding repositories respectively other than the Linux repository.
* For each repository {R1, R2, R3} find out if Cl1 is the core contributor for any of them and similarly we should find out the same for {R4, R5, R6}.
* Assuming that R1, R2 have Cl1 as the core-contributor satisfying the core contributor rule for each of them. We can now change the terminology of Cl1 to Cl1{R1, R2}.
* Similarly, we can assume that Cl2 can be changed as Cl2{R5, R6}.

**Note:** Note that Cl1{R1, R2} is the core contributor of R1, R2 and only a contributor of Linux repository and the same is applicable to Cl2{R5, R6}.

* Now, we know that Cl1{R1, R2} has already forked Linux repository(since it is the contributor of Linux already) but in order to draw a citation link between R1, R2 and Linux repositories we should check if the fork to Linux has been done while Cl1{R1,R2} has been actively working on R1, R2.
* Let us assume that Cl{R1, R2} has forked Linux while he was actively working on R1 and not while working on R2. Similarly R6 satisfies the same condition for Cl{R5, R6}.
* Hence, we can draw a citation link from Linux to R2 and R6 and conclude that R2, R6 have cited Linux repository.

R2

Linux

R6

Fig2: Part B

**The above figure is the citation network formed for the given repositories**

**Part C**

From Part B we have R2 and R6 citing Linux repository. Now we should find out if there are any repositories from the list of repositories citing Linux.

* Based on our citation link rule we can assume that the core contributor of R2 has forked X (reference repository of linux) during the active period on R2.
* It implied that we can draw a citation link between R2 and X (R2 has cited X)

X

R2

Linux

Fig3:Part C

**Part D**

From Part B we have R5’s core contributor forked the reference Y and R3’s core contributor forked Z during their active period. Hence, we can draw a citation link between Y and R5 ; Z and R3 .

R5

Y

R3

Z

Final Citation network combining Part A, Part B and Part C

R5

R2

X

Z

Y

R6

Linux

R3

From the final citation network built for Linux repository:

References: X, Y, Z

Focal paper: Linux

Subsequent work: R2, R6

*ni*– *nj*

Disruption  =      ---------------

*ni*+ *nj*+ *nk*

In Linux Repository terms:

ni = Number of subsequent works citing Linux repository = 2

nj = Number of subsequent works citing both Linux repository and references of a Linux repository = 1

nk= Number of subsequent works citing references. = 2

D = 2-1/5 = 0.20