

**INTRODUCCIÓN A LOS MÉTODOS DEL
ANÁLISIS DE REDES SOCIALES.**

CAPÍTULO NOVENO

**Robert A. Hanneman. Departamento de Sociología de la
Universidad de California Riverside.**

NOTA PREVIA

Este documento esta traducido para la lista REDES con permiso del autor a partir de la versión electrónica disponible en <http://wizard.ucr.edu/~rhannema/networks/text/textindex.html> [Consulta: 20-02-02].

Estos capítulos han sido traducidos por [René Ríos](#), Departamento de Sociología de la Universidad Católica de Chile.

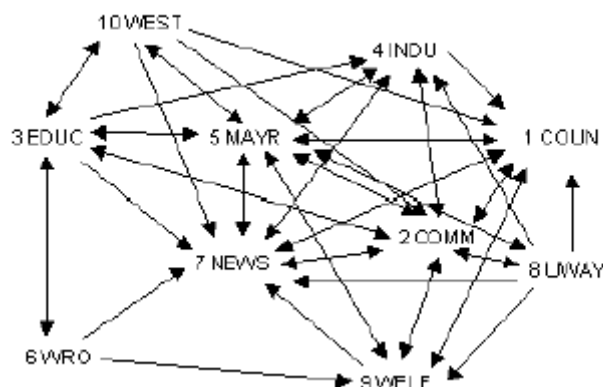
CAPÍTULO IX. MEDICIONES DE SIMILARIDAD Y EQUIVALENCIA ESTRUCTURAL

INTRODUCCIÓN

En esta sección vamos a examinar algunas de las maneras en que podemos definir y medir empíricamente el grado de equivalencia estructural o la similitud entre actores en razón de su posición en la red. En seguida vamos a examinar algunos de los posibles enfoques para analizar los patrones de equivalencia estructural basados en las medidas de similitud.

Recordemos que la equivalencia estructural significa que los actores son sustituibles. Esto es que los actores tienen un mismo patrón de relación con todos los otros. La equivalencia estructural exacta es rara en la mayoría de las estructuras sociales (como ejercicio, traten de imaginar como sería una estructura social en la que la mayoría de los actores fueren sustituibles por los demás). En consecuencia, a menudo calculamos medidas del grado en el que los actores son similares, y las usamos como base para identificar conjuntos de actores que son muy similares entre si y distintos de otros actores en otros conjuntos.

Para propósitos de ilustración, vamos a analizar los datos sobre envío y recepción de información entre burocracias de Knoke. Estos datos muestran vínculos dirigidos y están medidos a nivel nominal, esto es binario.



Este es un análisis bastante restringido, y no debe ser tomado muy en serio. Pero los datos son útiles para entregar una ilustración de los principales enfoques para medir la equivalencia estructural y para identificar actores que tienen similares posiciones reticulares.

MEDICIONES DE SIMILITUD ESTRUCTURAL

Podemos intentar establecer intuitivamente qué nodos son más similares entre sí con otros nodos mirando el diagrama. Notamos algunas cosas importantes. Parece ser que los actores #2, #5 y #7 pueden ser estructuralmente similares en tanto tienen lazos recíprocos y casi con todos los demás. Los actores #6, #8 y #10 son regularmente similares por que están relativamente aislados, pero no son estructuralmente similares porque están conectados a conjuntos distintos de actores. Más allá de esto, sin embargo, es realmente bastante difícil establecer la similitud estructural rigurosamente por medio de la inspección del diagrama.

Podemos ser más precisos si usamos la representación matricial. Esto además nos permite usar el computador para realizar el tedioso trabajo involucrado en el cálculo de índices para establecer la similitud. Se reproduce la matriz original de datos más abajo. Muchos de los rasgos que aparecían en el diagrama son también fácilmente captables en la matriz. Si miramos en las filas y contamos los grados de salida y si examinamos las columnas y contamos los grados de entrada podemos ver quienes son actores centrales y quienes aislados. Pero, de manera más general, podemos ver que dos actores son estructuralmente equivalentes en la medida que su perfil de puntajes en sus filas y columnas son similares.

	1 Coun	2 Comm	3 Educ	4 Indu	5 Mayr	6 WRO	7 News	8 UWay	9 Welf	10 West
1 Coun	---	1	0	0	1	0	1	0	1	0
2 Comm	1	---	1	1	1	0	1	1	1	0
3 Educ	0	1	---	1	1	1	1	0	0	1
4 Indu	1	1	0	---	1	0	1	0	0	0
5 Mayr	1	1	1	1	---	0	1	1	1	1
6 WRO	0	0	1	0	0	---	1	0	1	0
7 News	0	1	0	1	1	0	---	0	0	0
8 UWay	1	1	0	1	1	0	1	---	1	0
9 Welf	0	1	0	0	1	0	1	0	---	0
10 West	1	1	1	0	1	0	1	0	0	---

Se dice que dos actores son estructuralmente equivalentes si tienen similares patrones de vínculos con otros actores. Esto significa que las entradas en las filas y columnas de cada actor son idénticas entre sí. Si la matriz fuese simétrica, sólo necesitaríamos examinar pares de filas o columnas. Pero como estos datos son lazos dirigidos, debemos examinar la similaridad de los vínculos de emisión y de recepción (por supuesto, podemos estar interesados en la equivalencia estructural de sólo los lazos emisores o de los receptores). Podemos ver la similitud de los actores si expandimos la matriz listando un vector de fila seguido de un vector de columna de cada actor como columna, así:

1 Coun	2 Comm	3 Educ	4 Indu	5 Mayr	6 WRO	7 News	8 UWay	9 Welf	10 West
---	1	0	1	1	0	0	1	0	1
1	---	1	1	1	0	1	1	1	1
0	1	---	0	1	1	0	0	0	1
0	1	1	---	1	0	1	1	0	0
1	1	1	1	---	0	1	1	1	1
0	0	1	0	0	---	0	0	0	0
1	1	1	1	1	1	---	1	1	1
0	1	0	0	1	0	0	---	0	0
1	1	0	0	1	1	0	1	---	0
0	0	1	0	1	0	0	0	0	---
---	1	0	0	1	0	1	0	1	0
1	---	1	1	1	0	1	1	1	0
0	1	---	1	1	1	1	0	0	1
1	1	0	---	1	0	1	0	0	0
1	1	1	1	---	0	1	1	1	1
0	0	1	0	0	---	1	0	1	0
0	1	0	1	1	0	---	0	0	0
1	1	0	1	1	0	1	---	1	0
0	1	0	0	1	0	1	0	---	0
1	1	1	0	1	0	1	0	0	---

Para ser estructuralmente equivalentes, dos actores deben tener los mismos vínculos con los mismos otros actores - esto es, que las entradas en sus columnas deben ser idénticas (con la excepción de los autovínculos, que probablemente ignoraríamos en la mayoría de los casos). Habiendo organizado los datos de esta manera, hay algunos aspectos bastante obvios que podríamos hacer para proveer un índice que resuma cuán cercano a la equivalencia estructural está cada par de actores. Con algo de imaginación se puede llegar a otras ideas, pero los cuatro enfoques más comunes para indexar la equivalencia estructural son la correlación, las distancias euclidianas

cuadradas, *matches* (aciertos) y aciertos positivos. Más abajo, cada uno de estos resultados se muestra como una matriz de similitud o distancia.

COEFICIENTES DE CORRELACIÓN PEARSON.

La medida de similitud por correlación es particularmente útil cuando los datos de los lazos están valorados, esto es contienen información acerca de la fortaleza del vínculo y no sólo de su presencia o ausencia. Las correlaciones Pearson van entre -1.00 (que significa que dos actores tiene exactamente los mismo lazos opuestos entre sí), a través de cero (que implica que el conocimiento de los vínculos de un actor a un tercero no nos ayuda a estimar los lazos de otro actor con ese tercero) a +1 (implicando que los dos actores tienen exactamente el mismo par de vínculos con otros, una equivalencia estructural perfecta). Las correlaciones Pearson se usan a menudo para sumarizar equivalencia estructural pareada, porque el estadístico "r" se usa frecuentemente en estadística social. Si los datos no son verdaderamente nominales o si la densidad es muy alta o muy baja, las correlaciones pueden ser problemáticas y deberían examinarse los aciertos. Diferentes estadísticos usualmente entregan las mismas respuestas. A continuación se presentan las correlaciones Pearson entre los vectores concatenados de filas y columnas calculados con UCINET.

```

1 2 3 4 5 6 7 8 9 10
COUN COMM EDUC INDU MAYR WRO NEWS UWAY WELF WEST
-----
1 1.00
2 0.42 1.00
3 0.10 -0.48 1.00
4 0.50 0.42 0.13 1.00
5 0.33 0.79 -0.38 0.33 1.00
6 -0.07 0.15 0.00 -0.07 0.00 1.00
7 0.40 0.29 0.22 0.42 0.10 -0.37 1.00
8 0.63 0.37 0.26 0.63 0.29 -0.00 0.05 1.00
9 0.63 0.04 0.26 0.38 -0.10 -0.10 0.59 0.49 1.00
10 0.26 0.29 0.40 0.52 0.25 0.36 -0.03 0.38 0.13 1.00

```

Nota: Sólo se necesita la mitad de la matriz, dado que la similaridad XY es la misma que la YX. Hay un rango amplio de similaridades en la matriz. Los actores 2 y 5 son los más similares ($r=.79$); los actores 3 y 5, los más disimilares ($r=-.38$). La mayoría de los números son positivos, y muchos son sustanciales. Este resultado es consistente con la relativamente alta densidad y reciprocidad en estos datos, y es un resultado de ellas. Cuando las densidades son muy bajas, y los lazos no son recíprocos, las correlaciones pueden ser muy pequeñas.

Una manera de llegar a una síntesis aún más clara se obtiene al realizar un análisis de aglomerados (cluster) de la matriz de similitud. Lo que éste hace es agrupar juntos a los nodos que son más similares primero, en este caso los actores 2 y 5. Tomando los puntajes de cada conglomerado (cluster) que están más cerca de alguno de los otros conglomerados (método de vínculo único o vecino más cercano), se recalculan las similitudes, y en seguida se agrega el siguiente par más similar (podría ser otro par de actores, o el par 2 y 5 con algún otro actor). Este proceso continúa hasta que todos los actores estén unidos. A continuación se presenta un sumario de la matriz de similitud o proximidad, generado por un análisis de aglomeración de vínculo único.

HIERARCHICAL CLUSTERING OF CORRELATION MATRIX

```

1
Level 5 2 4 1 8 7 9 6 3 0
-----
0.787 XXX . . . . .
0.630 XXX . XXX . . . .
0.595 XXX . XXX XXX . .
0.587 XXX XXXXX XXX . .
0.409 XXX XXXXXXXXX . .
0.405 XXX XXXXXXXXX . XXX
0.242 XXX XXXXXXXXX XXXXX
0.140 XXXXXXXXXXXXXXX XXXXX
0.101 XXXXXXXXXXXXXXX

```

Notas: El resultado establece que los actores #5 y #2 son más similares (a nivel de .787 usando el método particular de aglomeración escogido). Este par permanece separado de todos los demás hasta bien avanzado el proceso de aglomeración. El siguiente par más similar es el constituido por el actor #1 y el #8 (a .630); luego el par #7, #9 (a .595). A una similaridad de .587 un grupo más grande es creado, compuesto por los actores #4, #1 y #8; el proceso continúa hasta que todos los actores se aglomeran. Notar que los actores #6, #3 y #10 se aglomeran en un grupo suelto, y están bastante distante de los restantes actores.

¿Cuántos grupos de nodos estructuralmente equivalentes hay en esta red? No hay una respuesta correcta. La teoría puede llevar a predecir que los flujos de información siempre producen tres grupos; o que los grupos siempre se dividen en dos facciones, etc. Si tenemos una teoría acerca de cuantos conjuntos de actores estructuralmente equivalentes debiera haber, podemos evaluar su bondad de ajuste con los datos. Más a menudo, no tenemos una teoría muy fuerte acerca de cuántos conjuntos de actores estructuralmente equivalentes podrían haber. En este caso, los resultados proveen algunas guías.

Nota que en un alto nivel de similaridad (.787), los diez actores se dividen en 9 grupos (5 y 2, y cada uno de los demás). Si permitimos que los miembros de los grupos sigan siendo equivalentes, aún si son similares sólo a nivel .60 (.585), podemos llegar a un cuadro bastante distinto: hay 6 grupos entre los diez actores. Si bajamos al nivel .40 de similitud, hay tres grupos con un excluido (el casi aislado WRO, # 6). Podríamos dibujar un diagrama del número de grupos (en el eje Y) contra el grado de similaridad al aglomerar (en el eje X). Este diagrama es uno del tipo de retornos marginales (llamado en análisis factorial un *Plot Screen*). su punto de inflexión es el más eficiente número de grupos. Más usualmente, uno busca una imagen del número de grupos estructuralmente equivalentes que es suficientemente simple de comprender, a la vez que sea suficientemente fuerte estadísticamente (con una relativamente alta similitud) para ser defendible.

El coeficiente de correlación Pearson, da considerable ponderación o peso a las diferencias más grandes entre puntajes particulares del perfil de actores, por la forma de cálculo, que eleva al cuadrado la diferencia entre puntajes entre vectores. Esto lo hace sensible a los valores extremos (en datos valorados) y a algunos errores. La correlación también mide la asociación lineal, y en algunos casos ésta puede ser una noción restrictiva.

DISTANCIAS EUCLIDIANAS

Una medida relacionada pero menos sensibles es la distancia Euclidiana. Esta es una medida de dis-similitud, pues es la raíz de la suma de las diferencias al cuadrado entre los vectores de los actores, es decir, las columnas de la matriz de adyacencia. En muchos casos, los análisis de distancias euclidianas y de correlación pearson arrojan el mismo resultados sustantivo. No obstante es una práctica recomendable analizar ambas.

He aquí los resultados basados en las distancias euclidianas.

```

EUCLIDEAN DISTANCE MATRIX
1 2 3 4 5 6 7 8 9 10
COUN COMM EDUC INDU MAYR WRO NEWS UWAY WELF WEST
-----
1 0.00
2 2.45 0.00
3 2.65 3.32 0.00
4 2.00 2.45 2.65 0.00
5 2.65 1.00 3.16 2.65 0.00
6 3.00 3.32 2.83 3.00 3.46 0.00
7 2.24 2.24 2.45 2.24 2.45 3.46 0.00
8 1.73 2.65 2.45 1.73 2.83 2.83 2.83 0.00
9 1.73 3.00 2.45 2.24 3.16 2.83 2.00 2.00 0.00
10 2.45 2.83 2.24 2.00 3.00 2.24 3.00 2.24 2.65 0.00

```

Nota: El patrón es bastante similar. Los actores 2 y 5 que tienen la correlación más fuerte (una medida de similaridad) tienen la menor distancia (una medida de disimilaridad). A diferencia del coeficiente de correlación, el tamaño de la distancia euclidiana no puede interpretarse. Todas las distancias serán positivas (o al menos no negativas), y las ideas de asociación positiva, no asociación o asociación negativa, que tenemos con el coeficiente de Pearson, no se pueden usar con las distancias. Noten que el rango de los valores de las distancias euclidianas es relativamente menor que el de las correlaciones que vimos anteriormente. La distancia mayor no es a la menor tantas veces como la correlación mayor es a la menor. Esto resulta de la manera que se calculan, dando menos ponderación o peso a los casos extremos.

```

AGLOMERACIÓN JERÁRQUICA DE LA MATRIZ DE DISTANCIAS EUCLIDIANAS
Level 5 2 7 4 1 8 9 6 3 0
-----
1.000 XXX . . . . .
1.732 XXX . . XXX . . . .
1.821 XXX . XXXXXX . . . .
1.992 XXX . XXXXXXXX . . .
2.228 XXX XXXXXXXXXXXX . . .
2.236 XXX XXXXXXXXXXXX . XXX
2.434 XXX XXXXXXXXXXXX XXXXX
2.641 XXX XXXXXXXXXXXXXXXXXXXX
3.014 XXXXXXXXXXXXXXXXXXXXXXXX

```

Nota: En este caso, el análisis de las distancias euclidianas dan exactamente la misma impresión acerca de cuáles conjuntos son más equivalentes. Es habitualmente lo que se obtiene con datos binarios. Con datos valorados, los resultados pueden ser distintos entre correlaciones y distancias. La representación gráfica del análisis de conglomerados claramente sugiere tres agrupaciones de casos ({2,5}, {7,4,1,8,9}, {6,3,10}). También sugiere, de nuevo, que nuestros actores nucleares [5,2] tienen un alto grado de equivalencia estructural - y son más sustituibles entre sí.

PORCENTAJE DE COINCIDENCIA EXACTA

En algunos casos los vínculos que examinamos pueden estar medidos al nivel nominal. Cada par de actores puede tener una relación de uno de diversos tipos (codificados como "a", "b" "c" o "1", "2", "3"). Aplicar distancias euclidianas o correlaciones a ese tipo de datos puede ser erróneo. En vez, estamos interesados en el grado en el cual una relación para el actor X es una coincidencia exacta con la correspondiente relación del actor Y. Con datos binarios, el porcentajes de coincidencia exacta pregunta a través de todos los actores para los que podemos hacer comparaciones, ¿qué porcentaje de las veces X e Y tienen la misma relación con otro actor? Esta es una medida de equivalencia estructural para datos binarios dado que coincide mucho con nuestra noción del significado de la equivalencia estructural.

PORCENTAJE DE COINCIDENCIA EXACTA

1 2 3 4 5 6 7 8 9 10

COUN COMM EDUC INDU MAYR WRO NEWS UWAY WELF WEST

1	1.00									
2	0.63	1.00								
3	0.56	0.31	1.00							
4	0.75	0.63	0.56	1.00						
5	0.56	0.94	0.38	0.56	1.00					
6	0.44	0.31	0.50	0.44	0.25	1.00				
7	0.69	0.69	0.63	0.69	0.63	0.25	1.00			
8	0.81	0.56	0.63	0.81	0.50	0.50	0.50	1.00		
9	0.81	0.44	0.63	0.69	0.38	0.50	0.75	0.75	1.00	
10	0.63	0.50	0.69	0.75	0.44	0.69	0.44	0.69	0.56	1.00

Nota: Estos resultados muestran la similitud de otra manera pero que es bastante fácil de interpretar. El número .63 en la celda 2,1 significa que, comparando al actor #1 con el #2, tienen el mismo vínculo (presente o ausente) a otros actores el 63% de las veces. La medida es particularmente útil con medidas nominales multi-categoriales de relaciones; también provee una escala para datos binarios.

AGLOMERACION JERARQUICA DE LA MATRIZ DE COINCIDENCIA EXACTA.

```
1
Level 5 2 4 1 8 7 9 6 3 0
-----
0.938 XXX . . . . .
0.813 XXX . XXX . . . . .
0.792 XXX XXXXX . . . . .
0.750 XXX XXXXX XXX . . .
0.698 XXX XXXXXXXXXXX . . .
0.688 XXX XXXXXXXXXXX . XXX
0.625 XXX XXXXXXXXXXX XXXXX
0.554 XXX XXXXXXXXXXXXXXXXX
0.432 XXXXXXXXXXXXXXXXXXXXX
```

Notas: En este caso, el resultado que se obtiene usando coincidencias es bastante similar al de correlaciones y distancias, como suelen serlo en la práctica. La aglomeración si sugiere que el conjunto [2,5] es bastante diferente al de todos los demás actores. Las correlaciones y distancias euclidianas tendían a enfatizar la diferencia del conjunto [6,3,10] y a aminorar la particularidad única de [2,5].

COEFICIENTES DE JACCARD

En algunas redes las conexiones son bastante escasas. En efecto, si se están investigando relaciones de tipo personal en organizaciones grandes, los datos pueden tener muy baja densidad. Donde la densidad es muy baja, las medidas de coincidencias, correlaciones y distancias pueden mostrar escasa variación entre los actores y pueden dificultar el discernimiento de los conjuntos de equivalentes estructurales (por supuesto, en redes muy grandes, con baja densidad, puede haber en realidad muy bajos niveles de equivalencia estructural).

Un enfoque para resolver este problema consiste en calcular el número de veces que ambos actores reportan una relación (del mismo tipo de vínculo) a los mismos terceros actores como un porcentaje del número total de vínculos reportados. Esto es, ignoramos los casos en que ni X o Y están ligados a Z, e inquirimos, qué porcentaje en común tienen en relación a la totalidad de vínculos presente. Esta medida se llama porcentaje de coincidencias positivas (por UCINET), o el coeficiente Jaccard (en SPSS). He aquí el resultado obtenido de los datos:

Percent of Positive Matches (Jaccard coefficients)

1 2 3 4 5 6 7 8 9 10

COUN COMM EDUC INDU MAYR WRO NEWS UWAY WELF WEST

```

-----
1 1.00
2 0.54 1.00
3 0.46 0.31 1.00
4 0.60 0.54 0.42 1.00
5 0.50 0.93 0.38 0.50 1.00
6 0.18 0.27 0.11 0.18 0.25 1.00
7 0.58 0.64 0.54 0.55 0.60 0.08 1.00
8 0.67 0.46 0.50 0.67 0.43 0.20 0.38 1.00
9 0.67 0.36 0.50 0.55 0.33 0.11 0.64 0.56 1.00
10 0.40 0.43 0.44 0.60 0.36 0.38 0.31 0.50 0.36 1.00

```

HIERARCHICAL CLUSTERING OF JACCARD COEFFICIENT MATRIX

```

1
Level 6 2 5 3 4 1 8 7 9 0
-----
0.929 . XXX . . . . .
0.667 . XXX . . XXX . . .
0.644 . XXX . XXXXX . . .
0.636 . XXX . XXXXX XXX .
0.545 . XXX . XXXXXXXXXXX .
0.497 . XXX XXXXXXXXXXXXX .
0.435 . XXXXXXXXXXXXXXXXXXX .
0.403 . XXXXXXXXXXXXXXXXXXX
0.272 XXXXXXXXXXXXXXXXXXX

```

Nota: De nuevo el cuadro básico emerge, sugiriendo que no importa mucho qué estadístico se usa para indexar la similaridad estructural entre un par de actores - al menos para obtener el cuadro general. La aglomeración de estas distancias esta vez, enfatiza la singularidad del actor # 6, el cual es aún más singular por esta medición debido al número relativamente pequeño del total de relaciones que tiene, que resulta en un nivel de similitud aún menor debido a que se ignora la ausencia de lazos conjuntos. Cuando los datos son ralos, y hay muy sustanciales diferencias en los grados de los puntos, es una buena opción para datos nominales o binarios, usar el coeficiente de coincidencia positiva.

Con algo de inventiva, se puede pensar en algunas otras formas de indexar el grado de similitud estructural entre actores. Se puede usar la rutina "proximities" de SPSS, que ofrece una colección grande de medidas de similitud, junto con una buena discusión de ellas. La elección de la medida debe estar orientada por la noción conceptual de qué es lo importante acerca de la similaridad de dos perfiles de relación para los propósitos del análisis particular. A menudo, con franqueza, hace poca diferencia, pero esto no es suficiente para ignorar la cuestión.

DESCRIPCIÓN DE CONJUNTOS DE EQUIVALENCIA

ESTRUCTURAL: MODELOS E IMÁGENES DE BLOQUES

Los enfoques que hemos examinado hasta ahora procuran proveer sumarios de cuán similares o distintos son un par de actores usando la vara de la equivalencia estructural. Estas matrices de similitud o proximidad (y a veces, de su opuesto, distancia), proveen una completa descripción de parejas de estos aspectos de posiciones en las redes. Pero es difícil ver el bosque por medio de los árboles, las matrices de distancia y similitud a veces son tan grandes y densas como las matrices de adyacencia o de vínculos de las que se derivan.

Un enfoque muy útil para obtener el cuadro general consiste en aplicar el análisis de conglomerados para intentar discernir cuantos conjuntos de equivalencia estructural hay, y cuáles actores están en cada conjunto. Vimos antes varios ejemplos acerca de cómo se puede usar el análisis de conglomerados para presentar cuadros más simples y sintéticos de patrones de similitud o disimilaridad de la posición estructural de actores. El análisis de conglomerados está fuertemente afinado en comparaciones de pares y jerárquicas. También asume la unidimensionalidad subyacente en los datos de similitud. A menudo es útil considerar otros enfoques que tienen distintos sesgos. Uno de ellos es el escalamiento multidimensional, que no será discutido aquí, ya que raramente se usa para analizar datos de equivalencia estructural. Examinaremos tres enfoques más comunes: CONCOR, análisis de componentes principales y búsqueda de tabú.

La matriz de similitud y el análisis de conglomerados no dicen cuáles son las similitudes que hacen que los actores de un conjunto sean similares y cuáles diferencias hacen que los actores en un conjunto sean distintos a los actores en otro. Para comprender las bases de similitud y diferencias entre actores estructuralmente equivalente es pertinente el enfoque del modelo de bloques y la matriz de imagen que se basa en él. Ambas ideas se han explicado antes. Vamos a mirar ahora como nos pueden ayudar a entender los resultados del CONCOR y de la búsqueda de tabú.

CONCOR

Este es un enfoque que se ha usado hace tiempo. Aunque su algoritmo se ve actualmente como algo peculiar, la técnica generalmente produce resultados significativos. Comienza correlacionando cada par de actores. Cada fila de esta matriz de correlación actor por actor, se extrae y se correlaciona con cada una de las otras filas. En un sentido, el enfoque es equivalente a preguntarse ¿cuán similar es el vector de similitudes del actor X con el vector de similitudes del actor Y? El proceso se reitera una y otra vez. Eventualmente los elementos de la matriz de correlación iterada converge en un valor de +1 o -1.

CONCOR entonces divide los datos en dos conjuntos sobre la base de esas correlaciones. Luego, en cada conjunto (si tienen más de dos actores) el proceso se repite. Continúa hasta que todos los actores son separados (o hasta que perdemos interés en el proceso). El resultado es un árbol binario que da lugar a la partición final.

Para ilustrarlo pedimos a CONCOR nos muestre los grupos que mejor satisfacen esta propiedad cuando creemos que hay cuatro grupos. Todos los algoritmos de bloques requieren que tengamos una idea a priori acerca de cuántos grupos hay.

Matriz de bloques

	1	4	8	9	7	3	5	2	6	10
1	---	0	0	1	1	0	1	1	0	0
4	1	---	0	0	1	0	1	1	0	0
8	1	1	---	1	1	0	1	1	0	0
9	0	0	0	---	1	0	1	1	0	0
7	0	1	0	0	---	0	1	1	0	0
3	0	1	0	0	1	---	1	1	1	1
5	1	1	1	1	1	1	---	1	0	1
2	1	1	1	1	1	1	1	---	0	0
6	0	0	0	1	1	1	0	0	---	0
10	1	0	0	0	1	1	1	1	0	---

Nota: La matriz de bloques ha reagrupado las filas y columnas (e insertado líneas divisorias o "particiones") para intentar poner juntos a los actores que tienen filas y columnas similares (dada la restricción de forzar en cuatro grupos). La matriz de bloques nos permite ver las similitudes que tienen los cuatro grupos (y también que las agrupaciones no son perfectas). El primer grupo [1,4,8,9] no presenta un perfil consistente de envío y recepción de información entre sus integrantes, no son un clique. La relación de este grupo [1,4,8,9] con el próximo [7,3] es interesante. Todos los actores del primer grupo envían información al 7 pero no a 3. Solamente un actor del primer grupo recibe información de [7,3] y lo hace de ambos integrantes. Cada miembro del grupo [1,4,8,9] envía y recibe información de ambos miembros de [5,2]. Finalmente el conjunto [1,4,8,9] es similar pues no envían ni reciben información de [6,10]. Los actores 5 y 2 tienen perfiles casi idénticos de la comunicación recíproca con todos los miembros de cualquier otro grupo, con la excepción de [6,10]. Por último [6,10] está aislado, pero envía a ambos [7,3]; y comparten el hecho de que sólo [3] les recíproca.

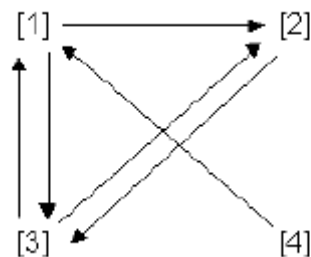
El diagrama de bloques permite ver con bastante claridad qué actores son estructuralmente equivalentes, y , más importante, qué de los patrones de sus relaciones define la similitud. En algunos casos desearíamos sintetizar la información aún mas, creando una imagen del diagrama de bloques. Para ello, combinamos elementos de las filas y columnas en grupos y caracterizamos las relaciones en la matriz como presente si despliegan una densidad mayor a la promedio (recordemos que en esta matriz la densidad general es .50). Si menos lazos están presente, se introduce un cero. La imagen de los bloques es:

	[1]	[2]	[3]	[4]
[1]	0	1	1	0
[2]	0	0	1	0
[3]	1	1	0	0
[4]	0	1	0	0

Nota: Esta "imagen" del diagrama de bloques tiene la virtud de una simplicidad extrema, aunque una buena cantidad de información acerca de las relaciones entre actores se ha perdido. Dice que los actores en el bloque 1 tienden a no enviar y recibir recíprocamente o con el bloque 4; tienden a enviar información a los bloques 2

y 3 pero reciben información sólo del 3. El segundo bloque de actores aparecen como receptores. Envían sólo al bloque 3, pero reciben información de cada uno de los otros grupos. El tercer bloque envía y recibe con los bloques 1 y 2, pero no con el 4. Finalmente, tal como antes, los actores del bloque 4 envían a los actores del bloque 2 pero no reciben información de ninguno de los otros bloques.

Se puede avanzar un paso más en la simplificación. Usando los conjuntos seccionados como nodos, y la imagen como un conjunto de adyacencias, podemos representar los resultados de la siguiente forma:



Comentario: Este gráfico más simplificado sugiere la marginalidad de 4, la reciprocidad y centralidad de las relaciones de 3 y la asimetría entre 1 y 2.

Hemos presentado los resultados de CONCOR en detalle para ilustrar la utilidad de permutar y generar bloques, calcular la imagen y graficar el resultado. CONCOR puede generar resultados peculiares que no son reproducidos por otros métodos. Si se usa CONCOR, es recomendable comparar sus resultados con otros algoritmos, una de esos métodos de comparación es la búsqueda tabú (que a su vez genera resultados peculiares que deben ser comparados).

La bondad de ajuste de un modelo de bloques se puede evaluar correlacionando la matriz permutada (el modelo de bloques) contra un modelo perfecto con los mismos bloques, esto es, uno en el que todos los elementos de un bloque son unos y todos los elementos de bloques cero son cero. Para un modelo CONCOR de dos divisiones (cuatro grupos) el r cuadrado (r^2) es .50, es decir, la mitad de la varianza en los lazos del model CONCOR se pueden explicar por un modelo estructural de bloques perfecto. Lo que puede ser considerado como aceptables aunque no sea una bondad muy buena ya que no hay criterios acerca de qué es una adecuada bondad de ajuste.

BÚSQUEDA DE TABÚ

Este método ha sido desarrollado recientemente y se basa en el uso intensivo de la computación. Usa un algoritmo más moderno y computacionalmente intensivo que el CONCOR, para tratar de implementar la misma idea de agrupar juntos a actores que son casi similares en un bloque. Lo hace buscando conjuntos de actores que, si se ponen en un bloque, producen la suma menor de varianzas de los perfiles de relación, dentro de los bloques. Si dos actores tienen relaciones similares las varianzas respecto al promedio del bloque será pequeña. De modo que la partición que minimiza la suma de varianzas dentro de los bloques, está minimizando la varianza general de los perfiles de relación. En principio, este método debería producir resultados similares (pero no necesariamente idénticos) que CONCOR. En la práctica no siempre es así. A continuación se presentan los resultados de este método con una especificación de 3 grupos.

Matriz de Bloques Adyacentes

Blocked Adjacency Matrix

	5	2	7	10	6	3	4	1	9	8
5	---	1	1	1	0	1	1	1	1	1
2	1	---	1	0	0	1	1	1	1	1
7	1	1	---	0	0	0	1	0	0	0
10	1	1	1	---	0	1	0	1	0	0
6	0	0	1	0	---	1	0	0	1	0
3	1	1	1	1	1	---	1	0	0	0
4	1	1	1	0	0	0	---	1	0	0
1	1	1	1	0	0	0	0	---	1	0
9	1	1	1	0	0	0	0	0	---	0
8	1	1	1	0	0	0	1	1	1	---

Comentario: Los bloques producidos son similares al resultado con CONCOR. Un conjunto de actores [1,4,8,9] es idéntico. El bloque CONCOR [7,3] se divide en los nuevos resultados, con el actor 7 agrupado con [5,2] y el actor 3 con el [6,10]. Examinando la matriz permutada y de bloques adyacentes, se puede ver que regiones

amplias de la matriz tienen ahora varianza mínima - las entradas en las celdas son las mismas. Esto es lo que el modelo de bloques intenta hacer: producir regiones lo más grande posibles, con homogeneidad de puntajes. Podemos interpretar directamente los resultados o producir una imagen de bloques:

	[1]	[2]	[3]
[1]	1	0	1
[2]	1	1	0
[3]	1	0	0

Comentario: Esta imagen sugiere que dos de los tres bloques es algo más solidario, en tanto que sus miembros envían y reciben información entre ellos (bloques 1 y 2). Los actores en el bloque 3 son similares por cuanto no intercambian información entre ellos directamente. Lo que estamos viendo es una especie de jerarquía entre bloques 1 y 2, con el bloque 3 más aislado. Estos patrones se pueden representar en un gráfico.



La bondad de ajuste de un modelo de bloques puede ser evaluado correlacionando la matriz permutada (el modelo de bloques) contra un modelo perfecto con los mismos bloques (uno en el que todos los elementos de un bloque son unos y todos los elementos de bloques cero son cero). Para el modelo de búsqueda tabú para tres grupos, este r^2 es .47. El que no es muy grande. Notamos que es una bondad de ajuste tan buena como la del modelo CONCOR con un grupo adicional. Por supuesto, uno puede ajustar modelos para varios números de bloques y graficar las estadísticas r^2 para decidir cual de ellos es el óptimo.

ANÁLISIS FACTORIAL, DE COMPONENTES PRINCIPALES

Todos los enfoques de equivalencia estructural se centran en la similaridad de patrones de relación de cada actor con los demás. Si esos patrones son similares, entonces los actores son estructuralmente equivalentes.

El análisis de conglomerados, búsqueda tabú y CONCOR, comparten en común el supuesto de que se busca una similaridad global o singular en una única dimensión. Esto, sin embargo, no garantiza que la matriz de similitud entre los perfiles de relación de los actores sea unidimensional. El análisis factorial y el escalamiento multidimensional de la matriz de similitud o de distancia relaja este supuesto. Estos métodos escalares sugieren que, subyacentes a las similitudes observadas, puede haber más de un aspecto o forma del patrón de equivalencia estructural y que los actores que son estructuralmente similares en un aspecto o dimensión pueden ser disimilares en otra.

La implementación de UCINET del enfoque de componentes principales busca la similitud en los perfiles de distancias de un actor a los demás. Los otros enfoques, que vimos antes, examinan los lazos desde y hacia los actores con datos direccionados. Por supuesto el análisis factorial puede ser aplicado a cualquier matriz rectangular simétrica, de modo que la opción de UCINET es una entre otros enfoques posibles.

Se calculan las correlaciones entre pares de filas en la matriz de distancia. Luego se realiza un análisis de componentes principales y las ponderaciones se rotan (la documentación de UCINET no describe el método usado para decidir cuántos factores retener, ni se describe el método de rotación). A continuación se presenta parte del resultado obtenido con la rutina CATIJ de UCINET.

CATIJ

Type of data: ADJACENCY
 Loadings cutoff: 0.60

CATIJ Matrix (length of optimal paths between nodes)

	1	2	3	4	5	6	7	8	9	0
	C	O	C	O	E	D	I	N	M	A
	W	R	N	E	U	W	W	E	W	E
1 COUN	0	1	2	2	1	3	1	2	1	2
2 COMM	1	0	1	1	1	2	1	1	1	2
3 EDUC	2	1	0	1	1	1	1	2	2	1
4 INDU	1	1	2	0	1	3	1	2	2	2
5 MAYR	1	1	1	1	0	2	1	1	1	1
6 WRO	3	2	1	2	2	0	1	3	1	2
7 NEWS	2	1	2	1	1	3	0	2	2	2
8 UWAY	1	1	2	1	1	3	1	0	1	2
9 WELF	2	1	2	2	1	3	1	2	0	2
10 WEST	1	1	1	2	1	2	1	2	2	0

Correlations Among Rows of CATIJ Matrix

	1	2	3	4	5	6	7	8	9	10
	COUN	COMM	EDUC	INDU	MAYR	WRO	NEWS	UWAY	WELF	WEST
1 COUN	1.00	0.65	-0.10	0.88	0.71	0.38	0.71	0.89	1.00	0.54
2 COMM	0.65	1.00	-0.45	0.65	0.65	0.06	0.60	0.82	0.60	0.38
3 EDUC	-0.10	-0.45	1.00	0.05	-0.29	0.37	0.29	-0.31	0.22	0.26
4 INDU	0.88	0.65	0.05	1.00	0.71	-0.16	0.87	0.89	0.88	0.79
5 MAYR	0.71	0.65	-0.29	0.71	1.00	0.00	0.71	0.80	0.71	0.38
6 WRO	0.38	0.06	0.37	-0.16	0.00	1.00	-0.00	-0.22	0.37	0.12
7 NEWS	0.71	0.60	0.29	0.87	0.71	-0.00	1.00	0.80	0.87	0.38
8 UWAY	0.89	0.82	-0.31	0.89	0.80	-0.22	0.80	1.00	0.80	0.32
9 WELF	1.00	0.60	0.22	0.88	0.71	0.37	0.87	0.80	1.00	0.68
10 WEST	0.54	0.38	0.26	0.79	0.38	0.12	0.38	0.32	0.68	1.00

Rotated Factor Loadings of Actors on Groups

	1	2	3
1 COUN	-0.90	-0.17	-0.37
2 COMM	-0.71	-0.53	-0.10
3 EDUC	-0.04	0.96	-0.19
4 INDU	-1.00	0.07	0.17
5 MAYR	-0.79	-0.36	-0.02
6 WRO	-0.02	0.16	-1.00
7 NEWS	-0.90	0.14	0.04
8 UWAY	-0.90	-0.38	0.14
9 WELF	-0.95	0.12	-0.33
10 WEST	-0.68	0.37	-0.06

Group 1:
 COUN COMM INDU MAYR NEWS UWAY WELF WEST
 Group 2:
 EDUC
 Group 3:
 WRO

Comentario: El análisis factorial identifica una clara y fuerte dimensión, y dos adicionales que son esencialmente internas. El segundo y tercer factor se deben a la particularidad de los patrones de las distancias de EDUC y a los patrones de las distancias de WRO. Estas distancias no se pueden explicar por los mismos patrones subyacentes de los otros, lo que sugiere que estas dos organizaciones no pueden ser priorizadas en las mismas dimensiones de similitud de que las otras.

Este resultado es bastante distinto al obtenido con los otros métodos. En parte se debe al uso de distancias en vez de adyacencias. La principal razón para la diferencias es, sin embargo, que los métodos factoriales (y de escalamiento multidimensional) buscan por dimensionalidades en los datos, en vez de suponer unidimensionalidad. Los resultados sugieren que las organizaciones 3 y 6 no son estructuralmente equivalentes a las otras y que sus lazos difieren en su significado o función de los vínculos entre las otras organizaciones. En un sentido, el resultado implica que el intercambio de información para estas dos organizaciones puede ser de un tipo de vinculación distinto que el de las demás organizaciones.

Si se mira a la ubicación de nodos en la primera dimensión, se puede ver que las organización están ordenadas de una manera similar a la obtenida con los otros resultados. Uno podría imaginar que seleccionando puntos de corte a lo largo de la primera dimensión se obtendrían agrupaciones agrupaciones {4,8,1,7,8}, {5}, {2,10}, {3,6}. Estas son bastante similares, si no idénticas, a las encontradas en los otros análisis.

Es sugerente el uso del análisis factorial para purificar las agrupaciones por medio de la extracción y remoción de dimensiones secundarias. Como casi todos los usos del análisis factorial, sin embargo, debería ser usado juiciosamente. Los resultados pueden ser similares, o muy distintos a los obtenidos con métodos de escalamiento unidimensional. Como siempre, no hay ninguno inherentemente correcto.

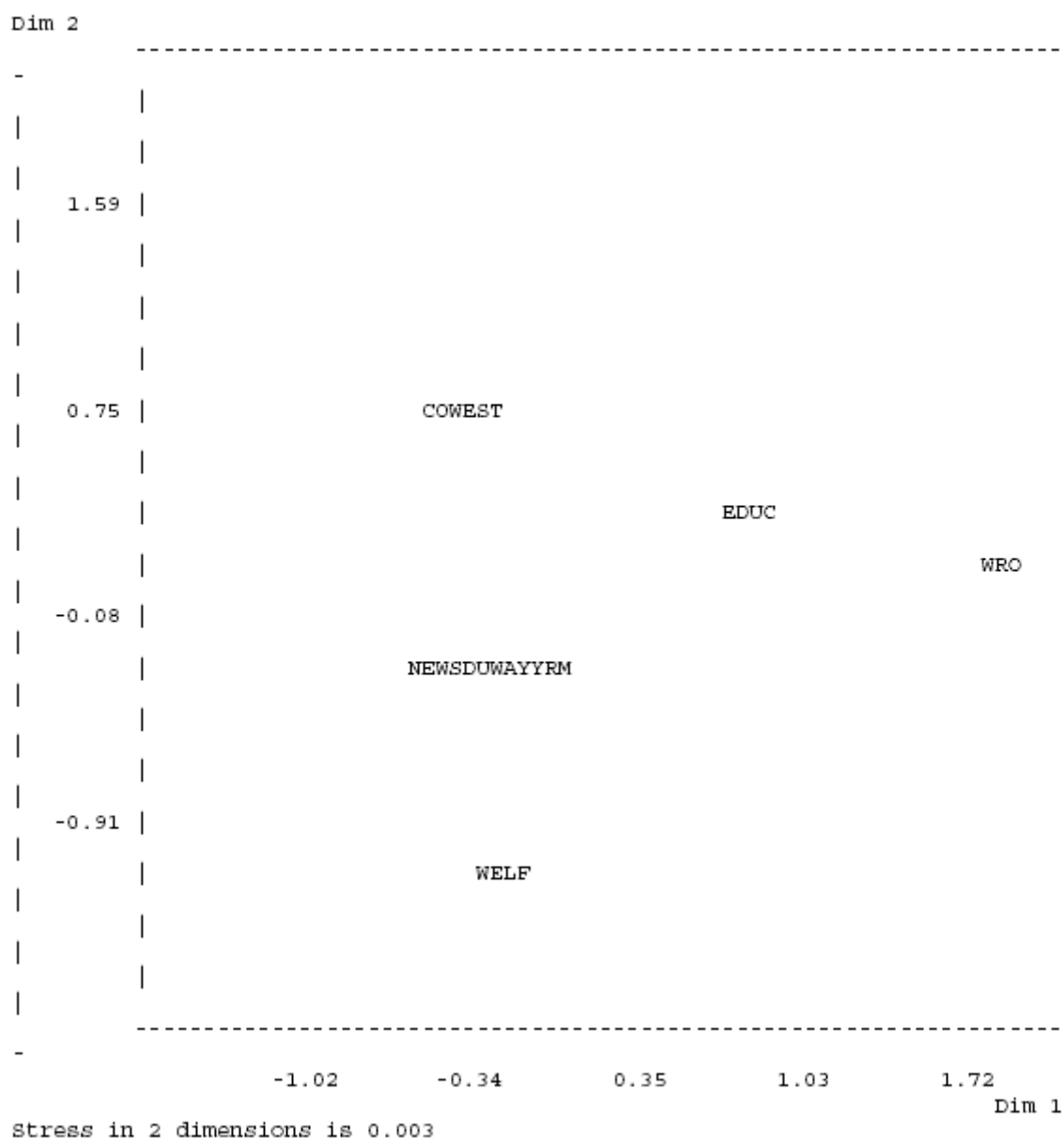
ESCALAMIENTO MULTIDIMENSIONAL

Una variación del algoritmo CATIJ es el escalamiento multidimensional, MDS, el cual, como el análisis factorial y de conglomerados, son una gran familia de técnicas algo diversas. Su mayor utilidad radica en que presenta los patrones de similitud o disimilaridad de los perfiles de relación entre actores (cuando se aplica a adyacencia o distancias), como un mapa de un espacio multidimensional. Este mapa permite ver cuán cercanos están los actores, si se aglomeran en el espacio multidimensional y

cuánta variación presentan en cada dimensión. A continuación se presentan los resultados de la rutina no paramétrica de UCINET usada para generar un mapa bidimensional de la matriz de adyacencia.

NON-METRIC MULTIDIMENSIONAL SCALING

	1	2
1 COUN	-0.49	0.94
2 COMM	-0.12	-0.24
3 EDUC	0.77	0.59
4 INDU	-0.42	-0.26
5 MAYR	-0.14	-0.24
6 WRO	1.91	0.23
7 NEWS	-0.58	-0.27
8 UWAY	-0.25	-0.25
9 WELF	-0.27	-1.47
10 WEST	-0.40	0.96



Comentario: El stress en dos dimensiones es bajo, esto es, la bondad de ajuste es alta. Esto sugiere que la tercera dimensión identificada por el análisis factorial previo puede no haber sido necesaria. EDUC y WRO son de nuevo identificados como estableciendo una dimensión distinta de las similitudes, en este caso dim 1. La aglomeración de casos en la dimensión principal, la dim 2, no corresponde muy cercanamente a los métodos de bloques ni factorial {1,10}, {2,4,5,7,8}, {9}.

Esto puede deberse a varias razones, tales como usar distancias euclidianas en vez de correlaciones, dos en vez de tres dimensiones, etc. Una vez mas no se trata de que una solución sea la correcta y las demás no. El enfoque MDS está mostrándonos

algo diferente y adicional acerca de las equivalencias estructurales aproximadas de estos puntos.

RESUMEN DEL CAPÍTULO IX

Hemos discutido la idea de equivalencia estructural de los actores y hemos visto algunas de las metodologías que son usadas para medirla, encontrar patrones en los datos empíricos y describir el conjunto de actores sustituibles.

La equivalencia estructural de dos actores es el grado al cual ambos tienen el mismo perfil de relaciones con otros en la misma red. La equivalencia exacta es rara en la mayoría de las estructuras sociales (una interpretación de equivalencia exacta es que representa una redundancia sistemática de los actores, que puede ser de alguna manera funcional para la red).

Aunque se puede a veces ver patrones de equivalencia estructural a ojo en una matriz de adyacencia o en un diagrama, casi siempre debemos usar métodos numéricos. Estos nos permiten lidiar con datos múltiples, grandes números de actores y datos valorizados, así como de tipo binario, como el que hemos examinado aquí.

El primer paso es producir una matriz de similitud o distancia para todo par de actores. Esta matriz sintetiza la similaridad o disimilaridad general de cada par de actores en términos de sus relaciones con otros. Hay varias maneras de calcular esos índices, los más comunes son la correlación Pearson, la distancia euclidiana, la proporción de coincidencias (para datos binarios) la proporción de coincidencias positivas (coeficiente Jaccard, también para datos binarios).

Se pueden usar varios métodos para identificar patrones en matrices de similitud o de distancia y para describirlos. El análisis de conglomerados agrupa juntos a los dos actores más similares entre sí, recalcula las similitudes e itera hasta que se ha combinado a todos los actores. Lo que produce una secuencia de aglomeración o mapa en el cual los actores se ubican en una jerarquía de inclusión incremental a grupos, que en consecuencia, son menos exactamente equivalentes. El escalamiento multidimensional y el análisis factorial se pueden usar para identificar qué aspectos de los perfiles de relación son más críticos para hacer a los actores más similares o distintos y también se pueden usar para identificar agrupaciones. La agrupación de actores estructuralmente equivalentes puede ser identificada por el método divisivo de iterar la matriz de correlación de actores (CONCOR) y por el método directo de

permutación y búsqueda de bloques de ceros y unos perfectos en la matriz de adyacencia (búsqueda tabú).

Una vez que se ha determinado el número de agrupaciones adecuado, los datos se pueden permutar y agrupar en bloques y calcular las imágenes. Estas técnicas hacen posible visualizar cómo son aproximadamente equivalentes los actores de un conjunto y por qué distintos conjuntos son diferentes. Nos permiten describir el significado de los grupos y ubicar a sus miembros en la red.

El análisis de equivalencia estructural a menudo produce hallazgos interesantes y reveladores acerca de los patrones de relación y las conexiones entre los actores individuales en una red. El concepto de equivalencia estructural busca operacionalizar la noción de que los actores pueden tener posiciones idénticas o casi iguales en una red, y, por lo tanto, ser directamente sustituibles por el otro. Una explicación alternativa es que los actores que son estructuralmente equivalentes enfrentan la misma matriz de restricciones y oportunidades en sus participaciones sociales.

El análisis sociológico no es acerca de sujetos individuales y el análisis estructural está ocupado, primariamente, con la idea más general y abstracta de roles o posiciones que definen la estructura del grupo, más que la ubicación de actores específicos en relación a otras personas. Para tal análisis, usaremos un conjunto de herramientas que permiten estudiar la replicación de subestructuras, la equivalencias automórfica, y los roles sociales, la equivalencia regular.