

COMPARACIÓN DE ALGORITMOS DE DETECCIÓN DE COMUNIDADES EN SISTEMAS ELÉCTRICOS DE POTENCIA

JOSE MORONTA R. ¹ Y CLAUDIO M. ROCCO S. ²

¹ Universidad Simón Bolívar. Departamento de Tecnología Industrial.

² Universidad Central de Venezuela. Facultad de Ingeniería.

e-mails: ¹ jmoronta@usb.ve, ² croccouv@gmail.com

Recibido: julio 2016

Aprobado para publicación: diciembre 2016

RESUMEN

Los sistemas eléctricos de potencia pueden ser modelados mediante conceptos derivados de la teoría de redes. De hecho, su estructura está basada en la presencia de subestaciones (nodos) interconectadas a través de elementos de transmisión (enlaces), tales como líneas de transmisión o transformadores. Una característica común en casi todas las redes es la presencia de comunidades, es decir, un grupo de nodos que están “relativamente, más densamente conectados entre sí pero escasamente conectados a otros grupos en la red”. El enfoque basado en la detección de comunidades ha sido utilizado en sistemas eléctricos de potencia para determinar qué nodos pueden ser agrupados, así como los enlaces entre las comunidades. Esta información podría ser útil para realizar otros estudios del sistema (por ejemplo, en un análisis de estabilidad dinámica, los enlaces inter-comunidades pudieran resultar los más sensibles ante eventos de falla, ya que estos conectan conjuntos de nodos que manejan grandes bloques de generación o de carga). El análisis de la literatura relacionada con la detección de comunidades en sistemas eléctricos de potencia, revela un factor común: los autores seleccionan un algoritmo específico para realizar la detección de comunidades, o comparan un algoritmo propuesto frente a otros ya existentes. Este artículo presenta la comparación simultánea de los resultados de varios algoritmos de detección de comunidades, aplicados exclusivamente en sistemas eléctricos de potencia, mediante la evaluación de la similitud de las particiones derivadas. Además, a diferencia de otras publicaciones que sólo consideran la información topológica de la red, este trabajo también compara las particiones derivadas en caso que la red posea información sobre características de los enlaces (por ejemplo, capacidad máxima, impedancia, entre otras). Para ilustrar la comparación de diversos algoritmos de detección de la comunidad, se analizan tres sistemas eléctricos de potencia.

Palabras clave: algoritmos de detección de comunidades, similitud de particiones, sistemas de potencia

COMPARISON OF COMMUNITY DETECTION ALGORITHMS IN ELECTRICAL POWER SYSTEMS

ABSTRACT

Electric power systems are modeled using concepts derived from network theory. Indeed, their structures are based on substations (nodes) interconnected through transmission elements (links), such as transmission lines or transformers. A common characteristic in almost all of the networks is the presence of communities, that is, a group of nodes that are “relatively, densely connected to each other but sparsely connected to other groups in the dense network”. Community detection in electric power systems allows determining what nodes are grouped as well as the inter-community links. This information is very important since it could be used as input to other power system evaluations (as an example, in dynamic stability assessments, inter-community links could be classified as the most important links in case of a failure since they connect groups of nodes that manage high load or generation blocks). The analysis of the literature regarding community detection in electric power systems reveals a common factor: the authors select a specific algorithm for community detection, or compare a new proposed algorithm only against other existing ones. This paper presents the simultaneous comparison of the results of several community detection algorithms, when considering only electric power systems, by evaluating the similarity of the partitions derived. In addition, unlike other publications that only consider the topological information of the network, this paper also compares the partitions derived when the network includes

information regarding characteristics of the links (e.g., capacity, impedance, among others). Three electric power systems are selected to illustrate the comparison of several community detection algorithms.

Keywords: community detection algorithms, partition's similarity, power systems

INTRODUCCIÓN

Una propiedad de las redes o grafos, en general, es la posible presencia de estructuras modulares o comunidades, definidas como un conjunto de nodos que están más densamente conectados entre ellos que con el resto de la red. Estas comunidades o grupos de nodos, corresponden a unidades funcionales dentro de las redes (Rocco et al. 2011), bien sea por poseer características similares o un desempeño similar en el comportamiento global de la red.

Los sistemas eléctricos de potencia pueden ser analizados desde la perspectiva de la teoría de redes o grafos, como se muestra, por ejemplo, en (Cruciti et al., 2004; Rossato et al., 2007 y Rocco, 2008). Al modelarse el sistema eléctrico como una red, pueden derivarse estudios o características típicas, tales como: conectividad, diámetro de la red, intermediación, algoritmos en redes (flujo máximo, caminos más cortos, entre otros), así como puede considerarse la presencia de comunidades. Ejemplos del uso del enfoque basado en la detección de comunidades en sistemas eléctricos de potencia, pueden encontrarse en Du et al., 2008; Rocco et al., 2011; Mei et al., 2011 y Schaub et al., 2012, entre otros.

Existen diversos algoritmos para la realizar la detección de comunidades en grafos. En general, los algoritmos pueden detectar particiones distintas (o grupos de comunidades), debido a los supuestos y características propias del algoritmo, el ámbito de aplicación de la red, a las características de los elementos de la red, entre otros factores. En cualquier caso, es posible comparar los resultados de las particiones obtenidas (Dannon et al., 2005; Lancichinetti et. al, 2009) a través de diferentes índices de similitud (Meilã, 2003; Dannon et al., 2005).

Las referencias mencionadas, presentan estudios de comparación de diversos algoritmos de detección de comunidades en forma general, esto es, mediante el uso de redes referenciales, mucha de ellas artificiales, en las que se conocen sus particiones.

Las investigaciones sobre comparaciones de algoritmos o diferentes enfoques para resolver un problema específico, es una actividad que ha sido realizada en diversas áreas, tales como pronóstico (Makridakis et al., 2000), exploración

de datos (data mining) (Borges et al., 2013), optimización evolutiva multiobjetivo (Coello et al., 2002), entre otras. Tal vez el ejemplo más didáctico de comparación de diferentes enfoques, se consigue en el área de pronóstico. Makridakis et al. (2000) decidieron hacer una comparación de 24 técnicas de pronóstico, mediante la evaluación de 3003 series de tiempo reales, con diversas características (por ejemplo, series estacionarias y/o estacionales, número de datos, presencia de valores atípicos, entre otras características). El resultado de la investigación fue la selección de una técnica que, “en promedio” produce buenos resultados, independientemente de la serie de tiempo. Sin embargo, como afirman los autores, esto no quiere decir que, para series de tiempo con características particulares, el método seleccionado sea el mejor.

La literatura específica relacionada con la detección de comunidades en sistemas eléctricos de potencia se ha centrado en el uso de un algoritmo en particular o solamente en la comparación de un nuevo algoritmo propuesto contra otros. Por ejemplo, en (Du et al., 2008), así como en (Schaub et al., 2012), se aplican algoritmos de detección de comunidades a sistemas eléctricos de potencia, pero sólo para comparar los modelos propuestos frente a modelos clásicos. En (Rocco et al., 2011), los autores seleccionan el método de detección de comunidades “Fast modularity” (Clauset et al., 2004) y lo aplican a diversas redes con el fin de evaluar la vulnerabilidad de la red en forma cualitativa.

Este trabajo presenta la comparación de un conjunto de algoritmos para detección de comunidades, bajo una óptica diferente: 1) la comparaciones se realizan en forma simultánea (*todos vs todos*) y solamente en topologías que representan exclusivamente sistemas de potencia, mediante la evaluación de la similitud de las particiones obtenidas, a partir de índices existentes; 2) a diferencia de las propuestas de comparación que aparecen en la literatura (de acuerdo con nuestro ámbito de aplicación), en este trabajo se incluyen comparaciones de redes ponderadas, esto es, redes en las que los enlaces poseen un peso que representa una característica del sistema, típica en sistemas eléctricos de potencia.

La comparación de los algoritmos de detección de comunidades se ilustra con el estudio de dos sistemas eléctricos de potencia reales (el sistema eléctrico venezolano

y el sistema eléctrico italiano) y una red test de IEEE (la red de 118 nodos).

La estructura del trabajo es como sigue: la sección Detección de comunidades y comparación de particiones presenta algunas definiciones fundamentales asociadas a la detección de comunidades, los principales algoritmos y la comparación de particiones; la sección Metodología propuesta describe los pasos necesarios para la comparación y presenta las redes reales prueba ser evaluadas. Los resultados se presentan en la sección Analisis de resultados y las conclusiones y trabajos futuros en la sección Conclusiones.

DETECCIÓN DE COMUNIDADES Y COMPARACIÓN DE PARTICIONES

Detección de comunidades en redes

A continuación se presentan, un grupo de conceptos básicos relacionados con la detección de comunidades:

Grafo: un grafo $G(V, E)$, es una estructura definida por un conjunto de nodos o vértices V y un conjunto de enlaces E .

Grafo ponderado: un grafo ponderado $G^w = (V, E, W)$ consiste en un conjunto V de nodos (o vértices), un conjunto E de enlaces, y un conjunto de valores (pesos) W que son números reales asociados a los enlaces y representan alguna característica particular de los mismos.

Comunidad o estructura modular: dado un grafo $G^w = (V, E, W)$, una comunidad es un subgrafo $G'^w = (V', E', W')$, cuyos nodos presentan una fuerte cohesión estructural, que puede ser cuantificada de distintas maneras (Boccaletti et al. 2006).

Partición: una partición C es la división de una red o grafo en comunidades, considerando que cada vértice o nodo del grafo pertenecen a alguna comunidad. Al incrementarse el número de nodos de una red crecen las posibles particiones de forma exponencial (Aldecoa, 2012). Por esto, es fundamental definir la partición óptima para una red. Sin embargo, la determinación de una partición óptima está asociada a la definición de indicadores de la "calidad" de la partición seleccionada.

Modularidad: La modularidad es la medida más empleada para "para determinar la fortaleza de las comunidades encontradas por los métodos de detección de comunidades" (Newman, 2006). La modularidad Q mide la fracción de los enlaces de la red que conectan los nodos del mismo tipo (es decir, los enlaces dentro de la comunidad) menos el valor esperado de la misma cantidad en una red con igual número de comunidades, pero considerando conexiones al azar entre los nodos. Los valores que se acercan al valor máximo, $Q=1$, indican la posible presencia de estructuras modulares. En la práctica, los valores de modularidad superiores a 0,3 sugieren la presencia de estructuras modulares (Newman, 2006). En (Newman et al., 2004) se propone una expresión para cálculo de la modularidad en redes ponderadas, ec. (1):

$$Q^w = \frac{1}{2 * m} \sum_{ij} \left[P_{ij} - \frac{P_i P_j}{2m} \right] \delta(c_i, c_j) \quad (1)$$

donde:

m : la suma de todos los pesos de los enlaces de la red

P_{ij} : el peso de los enlaces entre los nodos i y j

P_i, P_j : los pesos de los nodos i y j , ec. (2).

$$P_i = \sum_j P_{ij} \quad (2)$$

$\delta(c_i, c_j)$: es 1 si $c_i = c_j$, y 0 en cualquier otro caso.

c_i : Comunidad a la cual es asignado el nodo i

c_j : Comunidad a la cual es asignado el nodo j

Enlaces intracomunidades: se denominan así a los enlaces de una red que pertenecen a una comunidad, es decir, su nodo de llegada y de salida pertenecen a un mismo grupo (ver Figura 1).

Enlaces intercomunidades: se denominan así a los enlaces de una red que conectan dos comunidades, es decir, su nodo de llegada pertenece a una comunidad diferente que su nodo de salida (ver Figura1).

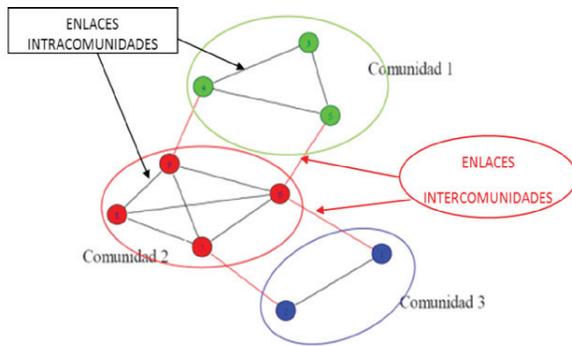


Figura 1. Red de ejemplo, con 3 comunidades y 4 enlaces intercomunidades

Existen diversos algoritmos para la realización de la detección de comunidades en grafos, por ejemplo, basados en la optimización del valor de la modularidad: FastGreedy (Clauset et al., 2004), Leading Eigenvector (Newman, 2006), Multilevel (Blondel et al., 2008), Infomap (Rosvall et al., 2007); algoritmos dinámicos: Spinglass (Reichardt et al., 2006), Walktrap (Pons et al., 2005); algoritmos divisivos: EdgeBetweenness (Girvan et al., 2002); alternativos: LabelPropagation (Raghavan et al., 2007), entre otros. Cada algoritmo posee características en cuanto a: complejidad computacional, requerimientos de datos, medidas para la validación, tipo de red que puede resolver (por ejemplo, dirigida, ponderada, entre otras). Para el lector interesado, además de las referencias de cada uno de los algoritmos, se puede consultar el artículo de (Fortunato, 2010), donde el autor presenta un estudio completo relacionado a los métodos de detección de comunidades. Los algoritmos de detección de comunidades utilizados en este trabajo se presentan en la Tabla 1 (todos estos algoritmos están codificados en el paquete igraph, bajo ambiente R (R Core Team, 2014)).

Tabla 1. Algoritmos de detección de comunidades evaluados

Método	Código	Método	Código
EdgeBetweenness	eb	Multilevel	ml
FastGreedy	fg	Leading Eigenvector	leig
WalkTrap	wt	Spinglass	sg
LabelPropagation	lp	Infomap	imc

Para efectos de presentación de los resultados, si el algoritmo es ejecutado sobre redes ponderadas, se incorporará el número 1 al final del código. Por ejemplo, si se aplica el método FastGreedy sobre una red no ponderada, los resultados se identificarán como “fg”. En cambio, si se aplica sobre una red ponderada “fg1”.

Comparación de particiones

El uso de diferentes algoritmos para la detección de comunidades en redes genera la necesidad de emplear índices de comparación, que permitan evaluar la similitud entre dos particiones obtenidas por distintos algoritmos. También puede considerarse, comparar particiones obtenidas a partir de la variación de alguna de las características de una misma red (por ejemplo, variando los pesos de los enlaces).

Existen dos grandes clasificaciones de las medidas para determinar similitud entre particiones:

Conteo de pares: donde se clasifican y ponderan los pares de nodos de acuerdo con su ubicación en las comunidades: si para dos particiones, dos nodos se encuentran juntos en una misma comunidad o separados, se consideran como un par de nodos positivos. En caso contrario, si en una partición el par de nodos está en una misma comunidad y en la otra están separados, este par de nodos se considera como un par negativo. Finalmente se totalizan los pares de nodos positivos y negativos, cada medida utiliza una fórmula distinta para calcular su valor de comparación, pero a partir de este conteo de pares. Basado en el conteo de pares encontramos: Rand Index (Rand, 1971), Adjusted Rand Index (ARI) (Hubert et al., 1985).

Teoría de la información: las medidas basadas en la teoría de la información tratan de evaluar cuánto se puede “adivinar” de una partición si ya se conoce la otra previamente. La información necesaria para hallar la segunda partición se determina mediante conceptos de Teoría de la Información como son la entropía o la información mutua (Cover et al., 1991; Strehl et al., 2002). Ejemplos de índices que utilizan este enfoque son: Variation Information (vi) (Meilă, 2003) y Normalized Mutual Information (nmi) (Dannon et al., 2005).

Estas medidas suelen normalizarse en el intervalo [0,1]: uno de los extremos significa que las particiones comparadas son independientes o totalmente distintas y el otro que las particiones comparadas son idénticas. Por ejemplo, para el Rand Index, el valor extremo 1 indica que las particiones son idénticas, mientras que para la medida Variation Information, el extremo cero identifica la igualdad de las particiones.

Sin embargo, en la literatura no existe un valor crítico o de referencia que permita concluir cuándo dos particiones pueden ser consideradas idénticas o totalmente distintas. En general, los autores fijan valores arbitrarios para estos valores críticos (Hric et al., 2014 y Marotta et al., 2015).

En este trabajo, se emplea el índice nmi , explicado a continuación. Sin embargo, cualquiera de los índices mencionados puede ser utilizado.

Normalized Mutual Information (nmi)

El índice, desarrollado en (Dannon et al., 2005), establece una medida basada en la definición de la matriz de comparación de particiones, donde las filas corresponden a las comunidades “reales” (o de la partición base) y las columnas a las comunidades “formadas” (o la partición de comparación). El elemento n_{ij} corresponde al número de nodos en la comunidad “real” i que aparece en la comunidad “formada” j .

El índice de comparación se establece de acuerdo con la Ec. (3):

$$nmi = I(C', C) = \frac{-2 \sum_{i=1}^{K_2} \sum_{j=1}^{K_1} n_{ij} \log \frac{n_{ij} * n}{n_i * n_j}}{\sum_{i=1}^{K_2} n_i \log \frac{n_i}{n} + \sum_{j=1}^{K_1} n_j \log \frac{n_j}{n}} \quad (3)$$

donde:

- K_1 = Número de comunidades formadas
- K_2 = Número de comunidades reales
- n_i = Suma de las filas i de la matriz de comparación de particiones
- n_j = Suma de las columnas j de la matriz de comparación de particiones

Si ambas particiones son idénticas, el índice es igual a 1.

METODOLOGÍA PROPUESTA

La comparación de algoritmos se realiza siguiendo los siguientes pasos:

1. Seleccionar el sistema eléctrico de potencia a estudiar.
2. Determinar la topología y las características a utilizar (con o sin pesos de los enlaces).
3. Determinar las particiones del grafo, mediante la aplicación de los algoritmos de detección de comunidades, codificados en la Tabla 1.
4. Determinar el índice de similitud entre todas las particiones obtenidas, así como: tiempo de cómputo, valor de modularidad, número de comunidades detectadas y número de enlaces intercomunidades.

En este trabajo se seleccionaron tres sistemas eléctricos de potencia: dos corresponden a sistemas reales: el sistema eléctrico venezolano y el sistema eléctrico italiano; mientras el tercero, es la red IEEE de 118 nodos.

Para cada uno de estos sistemas, se consideraron las versiones de red ponderada y no ponderada. Los valores asignados a los enlaces (para definir las correspondientes redes ponderadas) se generaron en forma aleatoria a partir de una distribución uniforme en $U[0,85; 0,95]$. Estos valores pueden representar las probabilidades de operación de los enlaces, o cualquiera otra característica normalizada (por ejemplo, longitudes, flujo de potencia reactiva, entre otras).

Como se mencionó anteriormente, el índice de similitud seleccionado es el índice nmi , aunque cualquiera de los índices mencionados puede ser utilizado.

Las evaluaciones de detección de comunidades e índice de similitud de particiones, se realizan mediante códigos desarrollados por los autores en el software libre R (R Core Team, 2014), plataforma: i386-pc-mingw32/i386 (32-bit), apoyado en la librería igraph 0.7.1 y ejecutados en un computador Intel® Pentium® Dual CPU T3200 @ 2.00 GHz, 2.00 GB de RAM.

A continuación se describen las características principales de los sistemas analizados.

Red Sistema Eléctrico Venezolano (VEN635)

Es una red de 635 nodos y 803 enlaces (Rocco, 2008), que representa el sistema eléctrico venezolano en un año particular (Figura 2a. Por razones de confidencialidad, el grafo se muestra sin ninguna referencia geográfica).

Red Sistema Eléctrico Italiano (ITAL127)

El sistema de potencia a evaluar corresponde a la red conformada por 127 nodos y 171 enlaces, asociada al sistema eléctrico italiano de 380kV (Rocco et al., 2011). El grafo asociado se observa en la Figura 2b.

Red Sistema Prueba IEEE 118 nodos (IEEE118)

Es una red de pruebas de la IEEE, compuesta por: 186 enlaces y 118 nodos (Power System Test Case, 2015). Esta red representa una parte del Sistema Eléctrico Norteamericano (específicamente en el medio Oeste de los Estados Unidos), en diciembre de 1962. La red asociada a este sistema puede observarse en la Figura 2c.

Los datos correspondientes a cualquiera de las redes pueden ser solicitados, por correo electrónico, a cualquiera de los autores.

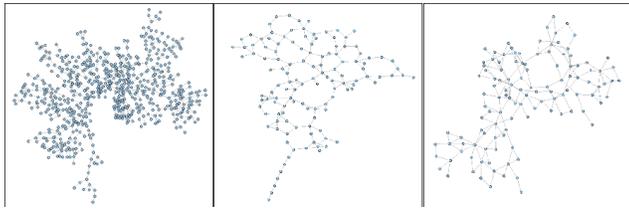


Figura 2. Grafos asociados a: Red Sistema Eléctrico Venezolano (izquierda); b: Red Sistema Eléctrico Italiano (centro); c: Red IEEE 118 nodos (derecha)

ANÁLISIS DE RESULTADOS

Red Sistema Eléctrico Venezolano (VEN635)

Para esta red, los resultados de los algoritmos, respecto al número de comunidades detectadas, presentan valores en el rango de 17 hasta 152 comunidades. La incorporación de pesos a los enlaces en esta red genera cambios en los resultados y, en general, la modularidad de las particiones disminuye ligeramente.

Los resultados de los atributos considerados, normalizados, se presentan de forma compacta en la Figura 3. Por ejemplo, los atributos obtenidos mediante el método Label Propagation, red ponderada, son distintos al resto de los métodos, en cuanto a comunidades y enlaces intercomunidades. Cabe señalar que el método Leadingeigenvector (leig1), para el caso de la red con pesos, no logra detectar comunidades, debido al tamaño de la red.

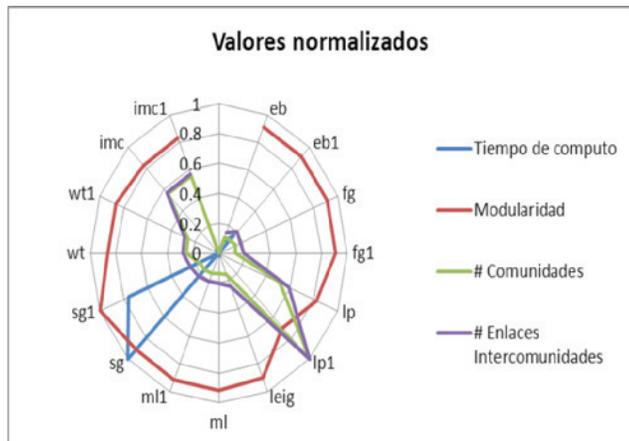


Figura 3. Resultados de algoritmos de detección de comunidades en la Red Sistema Eléctrico Venezolano

La comparación de las particiones obtenidas mediante el índice nmi , se ilustra cualitativamente, a través de un “heatmap” o mapa de calor, ver Figura 4. Las tonalidades más oscuras representan mayor similitud entre las particiones obtenidas por dichos métodos. El dendograma de la izquierda de la Figura 4 permite apreciar las relaciones entre los métodos de detección, es decir, agrupa los algoritmos con índices de similitud parecidos.

Como se mencionó anteriormente, los resultados obtenidos mediante el método Label Propagation son distintos al resto de los métodos al aplicarse sobre la red ponderada. Nótese como este método (primera línea del mapa de calor de la Figura 4) presenta la mayor cantidad de elementos en color blanco o gris claro, lo que indica valores bajos de nmi y con ello, que esta partición es la que más difiere del resto.

Por otra parte, el grupo de métodos ubicados en la esquina inferior derecha (leig, eb, eb1, sg1, fg1, ml1, sg, fg, ml) de la Figura 4, muestra zonas con tonalidades más oscuras, lo que sugiere que dichos métodos (LeadingEigenvector, EdgeBetweenness, Fastgreedy, Multilevel y Spinglass) presentan particiones similares, con índices nmi superiores a 0,85.

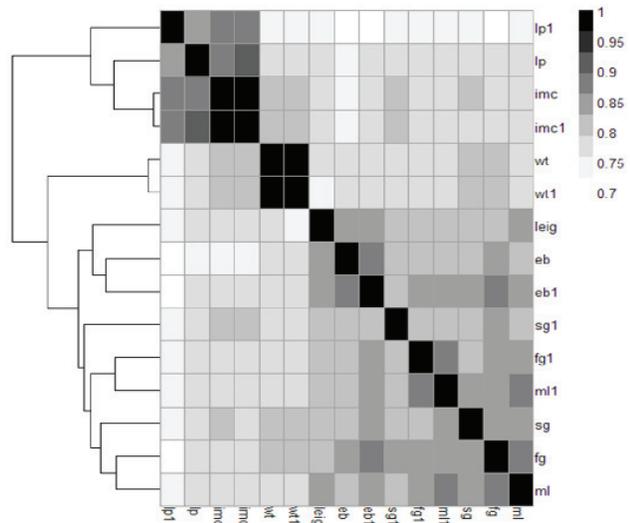


Figura 4. Comparación de particiones con Normalized Mutual Information del Sistema Eléctrico Venezolano

En contraste, la esquina superior derecha (lp1, lp, imc, imc1) presenta igualmente zonas de tonalidad oscura pero solo entre esos dos métodos, es decir, las particiones obtenidas por los métodos LabelPropagation e Infomap son similares entre sí, pero distintas al resto de los algoritmos (el resto de las celdas mostrados por estos métodos son de tonalidad más clara o incluso blancos).

Como se mencionó, el dendograma adjunto en la Figura 4 muestra, de forma global, la similitud entre las particiones obtenidas con los diferentes métodos. En este caso, la similitud se mide a través de la “distancia” que separa a dos o más métodos: a menor distancia, mayor similitud. Por ejemplo, los algoritmos (imc, imc1) o los algoritmos (wt, wt1) presentan alta similitud ente las particiones generadas ya que la distancia entre ellos es baja. Por otra parte, los métodos (fg1, ml1) o (fg, ml) también se agrupan pero a una distancia mayor. Finalmente, el grupo de algoritmos que se presentan en la esquina superior izquierda se une con el grupo formado por el resto de los algoritmos al nivel más alto del dendograma (esto es, a la máxima distancia), ratificando que las particiones que generan ambos grupos son distintas.

Red Sistema Eléctrico Italiano (ITAL127)

Los resultados normalizados para esta red se presentan de forma compacta en la Figura 5. El número de comunidades obtenidas son muy similares entre los algoritmos (varían desde 10 hasta 21), con excepción del algoritmo LabelPropagation sobre la red ponderada, que en esta red detecta 32 comunidades.

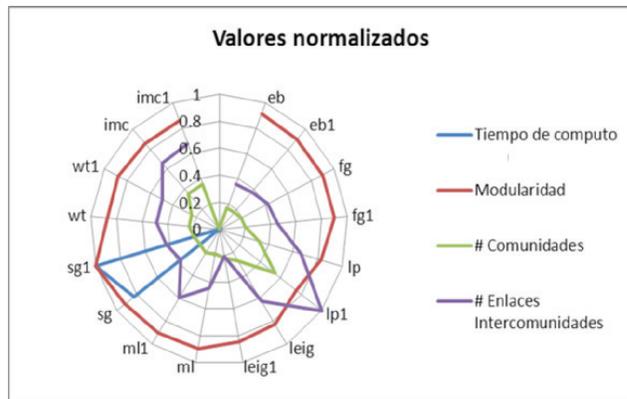


Figura 5. Resultados de algoritmos de detección de comunidades en el Sistema Eléctrico Italiano

En la Figura 6, se muestran las particiones obtenidas en la red italiana a partir del algoritmo EdgeBetweenness para la red sin ponderar y ponderada (eb y eb1). Se puede apreciar el efecto de la ponderación de los enlaces en la partición: los nodos y el enlace resaltados en mayor tamaño en la Figura 6, se separan de comunidad para el caso de la red ponderada, inclusive el enlace pasa de ser un enlace intracomunidad a uno intercomunidad.

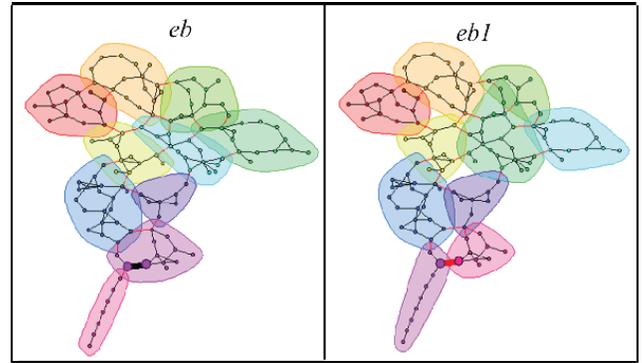


Figura 6. Particiones obtenidas sobre la red ponderada (der.) y sin ponderar (izq.) del Sistema Eléctrico Italiano, algoritmo EdgeBetweenness

A partir del mapa de calor de la Figura 7, se pueden identificar los métodos que presentan mayor similitud (zona donde se concentra la mayor tonalidad oscura del mapa): EdgeBetweenness, Fastgreedy y Spinglass (eb, fg y sg).

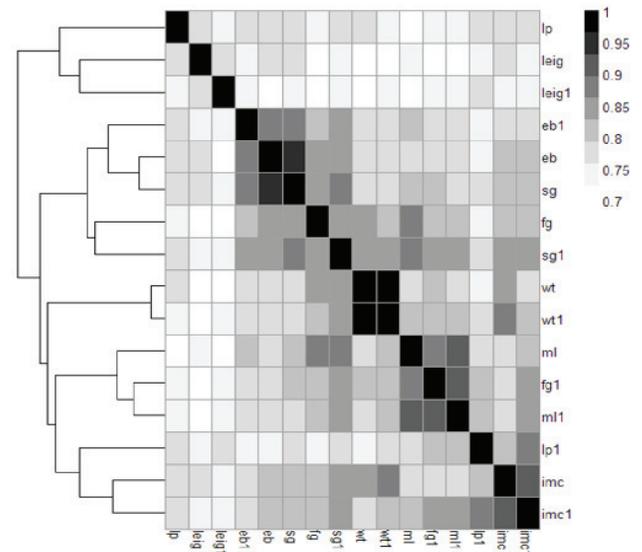


Figura 7. Comparación de particiones con Normalized Mutual Information del Sistema Eléctrico Italiano

En la Figura 7 las filas y columnas correspondientes al algoritmo LeadingEigenvector (sobre redes ponderadas y sin ponderar), poseen las tonalidades más claras lo cual indica que las particiones obtenidas con este método difieren del resto de los métodos. En general, el índice de similitud para el método LeadingEigenvector respecto al resto de algoritmos, está alrededor de 0,75.

Red Sistema Prueba IEEE 118 nodos (IEEE118)

Los resultados para la red IEEE, son consistentes con los obtenidos en las redes anteriormente analizadas. Los algoritmos LabelPropagation e Infomap, generan particiones con más comunidades y enlaces intercomunidades que el resto de los algoritmos (como se observa en el gráfico normalizado de la Figura 8).

El mayor tiempo de cómputo, en relación al resto de los algoritmos y en todas las redes, corresponde al método Spinglass. Para este caso particular, el Spinglass utilizó 6,47 y 6,86 segundos, para la red no ponderada y ponderada, respectivamente. Mientras el resto de los algoritmos presentan valores de tiempo de cómputo entre de 0,01 y 0,1 segundos.

La Figura 9 permite comparar, de forma cualitativa, dos particiones diferentes, obtenidas una por el algoritmo EdgeBetweeness y la otra por el algoritmo LabelPropagation (ambos aplicados sobre la red ponderada): la partición de la izquierda corresponde al algoritmo EdgeBetweeness y el de la derecha al LabelPropagation. La partición de la derecha tiene 27 comunidades versus ocho comunidades en la partición de la izquierda, y 65 enlaces intercomunidades versus 26 respectivamente. Al comparar estas particiones mediante el índice nmi , se obtiene uno de los valores más bajos 0,690, entre los algoritmos comparados para esta red.

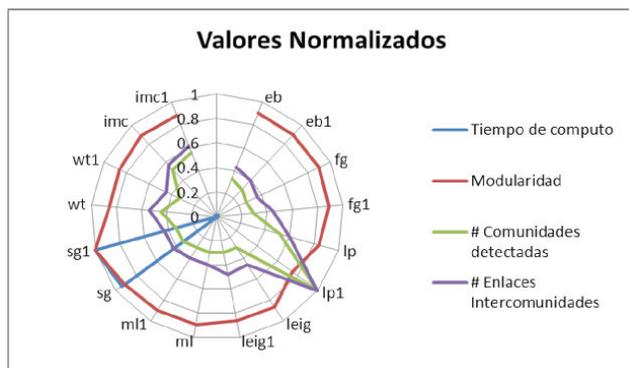


Figura 8. Resultados de algoritmos de detección de comunidades en el Sistema Prueba IEEE 118 nodos

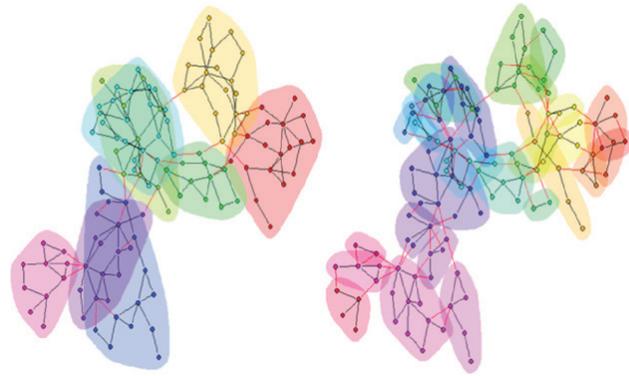


Figura 9. Particiones obtenidas sobre la red ponderada del Sistema Red Prueba IEEE 118 nodos, algoritmo EdgeBetweeness (izq.) y algoritmo LabelPropagation (der.)

CONCLUSIONES

Las redes eléctricas seleccionadas presentan estructuras modulares o comunitarias, con valores de modularidad superiores a 0,3, según el criterio establecido por (Newman, 2004). Las redes analizadas presentan topología descentralizada o, en el argot de los sistemas eléctricos, redes malladas. Por esto, la detección de comunidades tiene un mayor campo de aplicación en sistemas de transmisión y subtransmisión de energía eléctrica.

Los algoritmos empleados sobre las redes analizadas en este trabajo presentan bajos tiempos de cómputo (con la excepción del método Spinglass) y proporcional al número de nodos y enlaces (es decir, al tamaño de la red).

Los resultados obtenidos sugieren que los algoritmos EdgeBetweeness, FastGreedy y Spinglass son los más idóneos para la detección de comunidades en sistemas eléctricos de potencia, ya que presentan valores de modularidad e índices de comparación de particiones muy parecidos. Por otra parte, los algoritmos LeadingEigenvector e Infomap generan particiones que son poco consistentes con los resultados del resto de los algoritmos.

Cabe señalar que los índices de similitud utilizados en este trabajo permiten sólo una evaluación “global” de las particiones obtenidas por los algoritmos y supone que los pesos se conocen en forma precisa. En trabajos futuros se pretende evaluar el efecto de posibles incertidumbres en los pesos asociados a los enlaces de las redes y determinar la robustez de las particiones y comunidades bases (esto es, sin considerar incertidumbre).

REFERENCIAS

- ALDECOA R. (2012). *Detección de comunidades en redes complejas*. Tesis de Maestría no publicada. Universidad Politécnica de Valencia, Valencia, España.
- BLONDEL, V., GUILLAUME, J., LAMBIOTTE, R., & LEFEBVRE, E. (2008). Fast Unfolding of Communities in Large Networks. *Journal of Statistical Mechanics: Theory and Experiment*, (10); pp. 10008.
- BOCCALETTI S., LATORA V., MORENO Y., CHAVEZ M., & HWANG D.U. (2006). Complex Networks: Structure and Dynamics. *Physics Reports*. 424; pp. 175-308.
- BORGES L., NUNES P., MARQUES V., & BERNARDINO J. (2013). Comparison of Data Mining Techniques and Tools for Data Classification. *Proceedings of the International Conference on Computer Science and Software Engineering*; pp. 113-116.
- CLAUSET A., NEWMAN M. E., & MOORE C. (2004). Finding Community Structure in Very Large Networks. *Physical Review E*. 70(6); pp 066111.
- COELLO C., VAN VELDHUIZEN D., & LAMONT G. (2002). *Evolutionary Algorithms for Solving Multi-Objective Problems*. Kluwer Academic Publishers, New York.
- COVER T., & THOMAS J. (1991). *Elements of Information Theory*. John Wiley & Sons.
- CRUCITTI P., LATORA V., MARCHIORI M. (2004). A Topological Analysis of the Italian Electric Power Grid. *Physica A*. 338; pp. 92-97.
- DANON L., DIAZ-GUILERA A., DUCH J., & ARENAS, A. (2005). Comparing Community Structure Identification. *Journal of Statistical Mechanics: Theory and Experiment*. 09; pp. 09008.
- DU, N., WANG, B., & WU, B. (2008). Community Detection in Complex Networks. *Journal of Computer Science and Technology*, 23(4);, pp. 672-683, 2008.
- FORTUNATO S. (2010). Community Detection in Graphs. *Physics Reports*. 486; pp. 75-174.
- GIRVAN M., & NEWMAN M. E. (2002). Community Structure in Social and Biological Networks. *Proceedings of the National Academy of Sciences*. 99; pp. 7821-7826.
- HRIC D., DARST R. K., & FORTUNATO S. (2014). Community Detection in Networks: Structural Clusters Versus Ground Truth. *Physical Rev. E*. 9; pp. 062805.
- HUBERT L., & ARABIE P. (1985). Comparing Partitions. *Journal of classification*. 2(1); pp. 193-218.
- MAKRIDAKIS S., & HIBON M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting*; 16(4); pp. 451-476.
- MAROTTA L., MICCICHÈ S., FUJIWARA Y., IYETOMI H., AOYAMA H., GALLEGATI M., & MANTEGNA, R. N. (2014). Bank-Firm Credit Network in Japan: An Analysis of a Bipartite Network. e0123079.
- MEILĀ, M. (2003). Comparing Clusterings by the Variation of Information. *Learning Theory and Kernel Machines*. Springer Berlin Heidelberg. pp. 173-187.
- NEWMAN M. E. (2006). Modularity and Community Structure in Networks. *Proceedings of the National Academy of Sciences*. 103; pp. 8577-8582.
- NEWMAN M. E., & GIRVAN M. (2004). Finding and Evaluating Community Structure in Networks. *Physical review E*. 69(2); pp. 026113.
- NEWMAN, M. (2006). Finding Community Structure in Networks Using the Eigenvectors of Matrices. *Physical review E*. 74 (3); pp. 036104, 2006.
- LANCICHINETTI A., & FORTUNATO S. (2009). Community Detection Algorithms: A Comparative Analysis. *Physical review E*; 80(5); pp. 056117.
- PONS, P., LATAPY, M. (2005). Computing Communities in Large Networks Using Random Walks. In: *Computer and Information Sciences-ISCIS*., pp. 284-293, 2005.
- POWER SYSTEMS TEST CASE ARCHIVE: University of Washington (2015). Disponible en: <http://www.ee.washington.edu/research/pstca/>. Consultado: 02 de Junio de 2015.
- R CORE TEAM (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

- RAGHAVAN, U., ALBERT, R., & KUMARA, S. (2007). Near Linear Time Algorithm to Detect Community Structures in Large-scale Networks. *Physical Review E*. 76 (3); pp. 036106.
- RAND W. M. (1971). Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical association*. 66(336); pp. 846-850.
- REICHARDT, J., & BORNHOLDT, S. (2006). Statistical Mechanics of Community Detection. *Physical Review E*, 74(1); pp. 016110.
- ROCCO C. (2008). Análisis del sistema eléctrico venezolano desde la perspectiva de la teoría de redes complejas. *Revista de la Facultad de Ingeniería Universidad Central de Venezuela*. 23(1); pp. 103-109.
- ROCCO C., & RAMIREZ-MARQUEZ J. (2011). Vulnerability Metrics and Analysis for Communities in Complex Networks. *Reliability Engineering and System Safety*. 96; pp.1360–1366.
- ROSATO V., BOLOGNA S., TIRITICCO F. (2007). Topological Properties of High-voltage Electrical Transmission Networks. *Electric Power Systems Research*. 77; pp. 99-105.
- ROSVALL M., Y BERGSTROM C. (2008). Maps of Information Flow Reveal Community Structure in Complex Networks. *Proceedings of the National Academy of Sciences*. 105 (4); pp. 1118-1123.
- SCHAUB M., DELVENNE J., YALIRAKI S., & BARAHONA M. (2012). Markov Dynamics as a Zooming Lens for Multiscale Community Detection: Non Clique-like Communities and the Field-of-view Limit. *PloS one*. 7(2); pp. e32210.
- STREHL A., & GHOSH J. (2002). Cluster Ensembles a Knowledge Reuse Framework for Combining Multiple Partitions. *Journal of machine learning research*; 3; pp. 583-617.