Software de Genealogías

Jorge E. Miceli¹
Sergio Guerrero²
Universidad de Buenos Aires- Argentina

Teniendo como antecedente la elaboración de un trabajo analítico centrado en el vínculo parentesco-posesión de ganado en la reserva patagónica tehuelche del Chalía, en el sur argentino, tuvimos como meta la automatización de las complejas tareas de carga de información que antes habíamos ejecutado manualmente. El resultado de esta preocupación es el programa que presentamos, que consta básicamente de un módulo de carga de datos parentales, capaz de mapear en diagramas arbolados las relaciones de descendencia que ligan a los miembros de una comunidad, y un módulo de generación de una matriz relacional procesable directamente por programas de ARS como el UCINET y PAJEK. Las diferentes representaciones generadas pueden almacenarse tanto en archivos de Excel como en archivos de texto con formato nativo directamente compatible con los programas nombrados, lo cual permite una extrapolación directa y sin intervención humana de la información ingresada por la interfase de carga. El desarrollo de esta aplicación nos otorga varias ventajas por sobre los métodos manuales de volcado de información. El proceso es mucho más veloz, consistente y confiable que apelando al ingreso manual, y además se pueden recorrer y modificar los datos de manera completamente intuitiva y clara, sin luchar con las dificultades inherentes a la interacción directa con la base de datos en su formato relacional. Además, la alimentación de información realizada por esta vía se desarrolla echando mano a la misma semántica del problema que se quiere analizar, cosa imposible de hacer si el usuario ejecuta la carga en la interfase del software final especializado en manejo de ARS.

1-Presentación y Funcionalidad general del programa

Con el objetivo de aplicar la teoría de análisis de redes a un conjunto de información etnográfica que resulte de nuestro interés, utilizamos el trabajo de campo hecho por Marcelo Muñiz en 1997 (transformado después en tesis de licenciatura) sobre la comunidad tehuelche del valle del Chalía, ubicada en el sudoeste de la provincia del Chubut, en la República Argentina.

Las características de la producción y reproducción de la comunidad del valle del Chalía están determinadas por varios factores. El más importante de ellos, desde nuestra perspectiva, es la especial relación económica establecida por el régimen de tenencia de la tierra y la condición étnica de

_

¹ jorgemiceli@hotmail.com

² sergue@fullzero.com.ar

los pobladores. La actual organización productiva de la colonia es el resultado de respuestas que esta comunidad fue elaborando históricamente en su conflictiva y asimétrica relación con la sociedad mayor, sobre la base de sus propios valores y tradiciones. Las relaciones de parentesco regulan el acceso de los productores a los medios de producción, cuya propiedad es comunitaria pero cuyo usufructo es individual. Al interior de la unidad doméstica organizan la inversión de fuerza de trabajo y su reproducción, mientras al exterior mantienen vínculos de reciprocidad con las restantes unidades, actuando como mecanismo de inclusión-exclusión de individuos y grupos familiares ajenos a la comunidad.

La propiedad de la tierra es comunal y su usufructo individual, de allí la importancia del mecanismo hereditario para acceder a los medios de producción, dado que la tierra no tiene valor de cambio. En el caso del ganado -mayoritariamente ovino- es distinto: tiene valor de cambio y tanto varones como mujeres reivindican igualdad de derechos por los animales. Los tehuelches basan su principal actividad económica en la cría de ganado menor con una modalidad extensiva.

El objetivo general que motivó el trabajo de investigación original (Muñiz 1998) consistió en analizar la dinámica productiva y reproductiva de la comunidad específicamente ubicada en lo que hoy es la Reserva Quilchamal.

El objetivo específico apuntó a indagar sobre la tenencia del ganado por parte de los grupos domésticos, su composición y la racionalidad en la utilización de la fuerza de trabajo dentro y fuera de la reserva.

Para cumplir con ellos, se utilizaron tres tipos de información:

- 1)Información proveniente de entrevistas colectivas e individuales en todas las unidades familiares, así como a personal sanitario, funcionarios municipales y estancieros vecinos.
- 2)Información proveniente de documentos de archivos vinculados a la historia de la comunidad.
- 3) Bibliografía pertinente al tema de investigación.

Partiendo de este marco y utilizando la información recogida por Muñiz en su investigación nuestro trabajo (Quiroga et al 1999), en cambio, se propuso objetivos diferentes a la investigación referenciada: desarrollamos un abordaje exploratorio y de alcance descriptivo. Intentamos indagar en el potencial analítico que puede tener la teoría de redes y sus metodologías a la hora de representar los vínculos parentales, para poder, posteriormente, profundizar sobre la composición global del sistema de parentesco, pues inferimos (y esta inferencia resultó después corroborada) que la posición de cada unidad doméstica en el sistema de parentesco general de la comunidad contiene vínculos multivariados que están determinando en algún grado la tenencia de ganado por cada unidad doméstica. Dicha correlación no era perfecta ni mucho menos, pero en general las unidades domésticas pertenecientes a grupos con índices de centralidad mas elevados en la red de parentesco poseían mayor cantidad de cabezas de ganado.

Teniendo como antecedente este primer análisis, nosotros tuvimos como meta la automatización de las complejas tareas de carga de información que antes habíamos ejecutado manualmente. El resultado de esta preocupación es el programa que presentamos, que consta básicamente de un módulo de carga de datos parentales, capaz de mapear en diagramas arbolados las relaciones de descendencia que ligan a los miembros de una comunidad, y un módulo de generación de una matriz relacional procesable directamente por programas de ARS como el UCINET y PAJEK.

Las diferentes representaciones generadas pueden almacenarse tanto en archivos de Excel como en archivos de texto con formato nativo directamente compatible con los programas nombrados, lo cual permite una extrapolación directa y sin intervención humana de la información ingresada por la interfase de carga.

El software corre bajo Windows desde su versión 95 en adelante, es fácilmente instalable y no requiere programas adicionales como soporte o complemento.

Esta Interfase permite ingresar información parental y visualizarla gráficamente de manera inmediata. El sistema tiene un nivel de profundidad de 3 generaciones como máximo, lo que no significa que de manera sucesiva no puedan representarse una cantidad mucho mayor de niveles genealógicos. La selección de niveles parentales es progresiva y secuencial, ya que los caminos de cada árbol no se despliegan todos a la vez. Si bien esto limita las posibilidades de representación simultánea de los datos, también permite una exploración mucho más focalizada de los lazos que conectan los diferentes núcleos familiares.

Por otro lado, la simbología usada es muy simple. Un triángulo representa un miembro masculino de la familia, y un círculo un miembro femenino. Al seleccionar una pareja cualquiera, aparecen en la caja inferior todos las alianzas que tienen como componentes a algún miembro del grupo de hijos de esa pareja. Para seguir seleccionando ramas del árbol se opera de la misma forma, y la visualización se produce en el siguiente nivel gráfico. (Ver Figura 1)

El programa genera una matriz parental directamente procesable por UCINET y PAJEK, dos programas muy usados en los análisis de Redes Sociales.

Se puede elegir el tipo de matriz a generar, o sea dicotomizada o no dicotomizada, o el criterio usado para considerar que 2 unidades domésticas están emparentadas. La primera opción establece el rango de datos que soportará cada celda de la grilla resultado. Si la matriz es dicotomizada cada celda cargará un 0 o un 1, es decir ausencia o presencia de relación sin especificar su grado. Si la matriz es no dicotomizada cada celda contendrá la cantidad exacta de vínculos entre 2 Unidades domésticas (Por ejemplo 5, 10 o 20).

La segunda opción ofrece la posibilidad de especificar cuando se considera que 2 UD están ligadas entre sí. Esta ligazón puede ser más restrictiva, si se consideran solo los padres en común, o menos restrictiva, si se consideran prácticamente todos los tipos de parentesco aplicables entre ambas.

El programa genera 2 tipos de archivos. El primer formato tiene extensión TXT (archivo de texto) y es procesable directamente por PAJEK y UCINET V. El segundo formato tiene extensión XLS (archivo de excel) y es procesable solo por UCINET V.

El nombre, la extensión y ubicación son asignados de manera automática por el programa. De existir un archivo en idéntica ubicación y con igual nombre el sistema pregunta si puede ser sobrescrito.

2-Aspectos epistemológicos y metodológicos generales del trabajo original

En el esquema que presentamos quisimos resumir la trayectoria general seguida para transformar los datos básicos de los cuales partimos en el comienzo de la investigación (una etnografía sobre la producción y reproducción económica en la comunidad tehuelche del Chalía) en información progresivamente más estructurada que nos permitiese responder los interrogantes específicos que nos fuimos haciendo a medida que avanzábamos.

Proponemos la existencia genérica de etapas o fases de procesamiento de la información en las que el volumen de datos existente, más algunas pautas ordenatorias y relacionales, producen una respuesta a determinadas preguntas planteadas. Estos interrogantes devienen tanto de nuestra plataforma teórica como de la aplicación de estrategias heurísticas ad hoc.

Las etapas propuestas estarían vinculadas por un conjunto de inferencias provenientes a su vez de la etapa anterior y se ajustan, excepto en el caso de la primera de ellas, al dominio de distintas aplicaciones del software disponible. El modo de articulación entre las etapas podría no ser exactamente el que se eligió, e incluso el orden seguido podría alterarse mediante la supresión o el cambio de lugar de algunas fases, pero nuestro propósito es aquí mucho más didáctico que prescriptivo y por ello nos interesa mucho más la exposición de la lógica global adoptada que la postulación de una manera unívoca de abordar este tipo de casos.

Nuestro trabajo se inició con una monografía construida con el preciso horizonte de una tesis de licenciatura, y este es el primer producto que tuvimos que analizar para generar información semánticamente significativa para las fases siguientes (FASE 1). Mas allá de la información relacional y de los nexos conceptuales presentes en esta fase, expresados de manera discursiva, el insumo de nuestra investigación fueron los datos genealógicos sobre posesión de ganado relevados en la investigación original.

El primer proceso de traducción necesario para pasar a la fase siguiente fue, entonces, la confección de una tabla de individuos que contuviese todos los atributos genealógicos y no genealógicos emergentes de la monografía original. Mediante la realización de este proceso pasamos de la información inestructurada de la monografía a la información tabulada en una base de datos ACCESS.

El volcado de datos en este nuevo formato fue posible por exclusiva intervención humana, pero luego de terminada esta fase tanto la elaboración de hipótesis provisionales como la confrontación con la información existente se ejecutó sobre el mismo conglomerado de información pero sometido a diferentes procedimientos taxonómicos y de cálculo. En la base ACCESS guardamos tanto información relacional (genealógica) como no relacional (posesión de ganado y otras), pero además tuvimos que desplegar una serie de consultas suplementarias para obtener de modo definitivo información sobre la cantidad de miembros que las diferentes Unidades domésticas compartían entre sí.

El segundo proceso de traducción o transformación se desarrolló transportando la información relacional final de la etapa anterior a una matriz inteligible para el programa UCINET V, y aquí tomamos otra decisión metodológica fundamental, que fue excluir la información cuantitativa total de los miembros que las diferentes UD comparten entre sí y tener en cuenta solo la existencia o ausencia de vínculos parentales. Este procedimiento, la dicotomización, implica la ganancia en calculabilidad dentro del contexto del UCINET V pero también la pérdida de granularidad o especificidad en la consideración de la información relevada. Debido a que la posibilidad de calcular índices globales de centralidad era para nosotros prioritaria,

preferimos no considerar la vinculación parental más allá de su expresión dicotómica, y por lo tanto fundar esta segunda traducción de ACCESS a UCINET en una exclusión de este nivel de información detallada con que en principio contábamos.

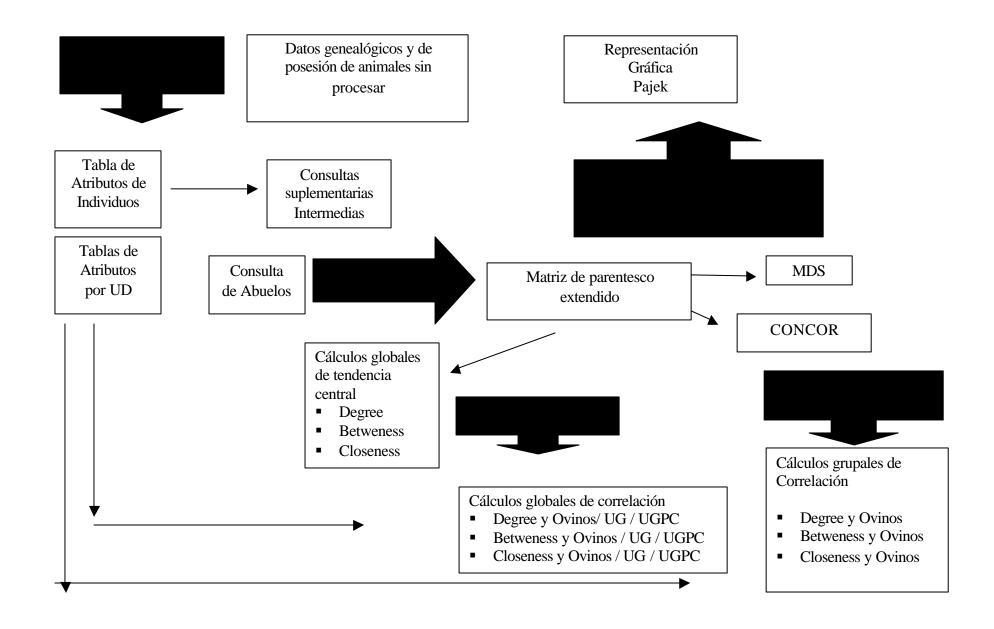
En el ámbito del UCINET V optamos por dos caminos distintos de traducción de información intraetapa: por un lado, los cálculos de las medidas globales de centralidad y por otro el cálculo de CONCOR y MDS como modos de representar espacialmente la vinculación parental de las UD. Ambas vías de indagación fueron tenidas en cuenta partiendo de potencialidades de cálculo presentes en el mismo software, pero nos posibilitaron arribar a conclusiones bastante definitivas sobre la existencia de grupos intracomunitarios que poseen más densa vinculación entre si que con el resto.

El tercer proceso de traducción de información se realizó también por prestaciones del software y sin ninguna acción restrictiva motivada metodológicamente. Simplemente se exportaron los archivos de la matriz de parentesco de UCINET y se procedió a la graficación en 2 y 3 dimensiones utilizando otro software llamado PAJEK.

El cuarto proceso de traducción es tal vez el más relevante desde el punto de vista de las conclusiones finales, porque se desarrolló en dos caminos simultáneos y obedeciendo a una intención verificatoria claramente dependiente de las hipótesis provisorias que pusimos en juego. En primera instancia tomamos los resultados de los cálculos globales de tendencia central y los cruzamos con la información sobre posesión de ganado a través de una serie de matrices genéricas que nos permitieron obtener conclusiones desalentadoras sobre la correlación global entre ambas variables. Sin embargo no sucedió lo mismo con los cálculos de correlación intergrupal, ya que estos, relevados íntegramente a través del uso de la planilla de cálculo Excel, nos permitieron verificar la existencia de fuertes correlaciones entre la centralidad parental y la posesión de animales.

Tanto los procesos de traducción o transformación de los datos que tienen lugar entre etapas diferentes como los que ocurren al interior de una etapa

no dependen, obviamente, de cuestiones exclusivamente técnicas, sino de una combinación particular de decisiones heurísticas y prestaciones puntuales que el software contiene. A pesar de ello, no podemos dejar de resaltar que la utilización de un soporte informático no es un mero instrumento ordenatorio del conjunto de datos con que contamos, sino, en muchos casos, una herramienta de fuerte incidencia en la obtención de conclusiones provisorias o definitivas que de otro modo sería mucho más engorroso generar. En nuestro caso particular los programas que utilizamos mostraron ser casi irremplazables para la presentación de los datos y para la ejecución de los cálculos elementales que debimos realizar. La creación de una matriz de parentesco global basada en las consultas de ACCESS, las representaciones gráficas hechas con PAJEK y la segmentación de grupos dispuesta por el CONCOR difícilmente hubiesen sido posibles en tiempo y forma si no hubiésemos echado mano del software en la medida en que lo hicimos, por más que la heurística investigativa y la articulación con el marco de análisis sean obra nuestra. En este marco, nosotros creemos que lo importante es que la interacción entre esta heurística y cada conjunto de resultados parciales ofrece el marco para sucesivas instancias de recorte del objeto de estudio. Estos recortes, por otro lado, no se sustentan en procedimientos reductores de carácter arbitrario, sino que aspiran a estar consistentemente apoyados por los resultados que el investigador considera pertinentes en cada caso.



3-Prestaciones del software en relación al trabajo original

El software que construimos se despliega complementariamente en relación a este esquema de trabajo. Más que como reemplazo de esta heurística, su utilidad puede verse a partir de los puntos de contacto que tiene en referencia a tal marco metodológico. Básicamente, el software de Genealogías automatiza las funciones que van desde lo que llamamos "Transformación 1" hasta la "Transformación 3". La "Transformación 1" es la que permite generar, a partir de datos no estructurados, un conjunto ordenado de referencias almacenadas tabularmente. El formato tabular es el que después permitirá el desarrollo de los diferentes cálculos estadísticos de orden grupal y global. La "Transformación 2", que es la que genera una matriz de parentesco extendido, está implícita y contenida en el software de Genealogías, ya que es la que hace posible la representación analítica y gráfica que son capaces de generar los programas UCINET y PAJEK. El único aspecto del tratamiento de datos que no hemos automatizado es el que permite el paso de la matriz de parentesco extendido hacia las "Transformaciones" 4 (1) y 4(2), en las cuales incluimos diferentes cálculos correlacionales globales y grupales. Debido a que la planilla de cálculo Excel ha demostrado prestaciones difíciles de imitar en la ejecución de estos procedimientos, no creemos que sea necesario llevar a un nuevo programa estas funciones.

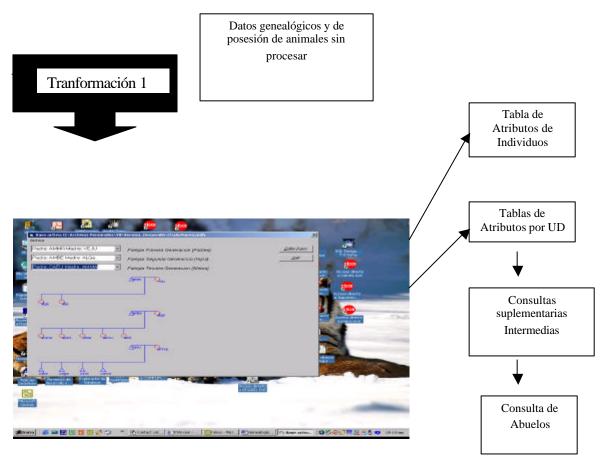


Fig. 1 La primera fase de nuestro esquema, que va de los datos inestructurados a los tabulares, estaría cubierta por nuestra interfase de carga de datos parentales.

La carga de datos se desarrolla a un nivel de detalle mayor que el de la familia. Cada individuo es portador de información adicional, como su nivel educativo, su edad o su nombre de pila, y esa información también puede ser asignada por uno de los menús del programa, reemplazando la simple carga de la tabla hecha en la base de datos usada como soporte. (Ver Fig.2)

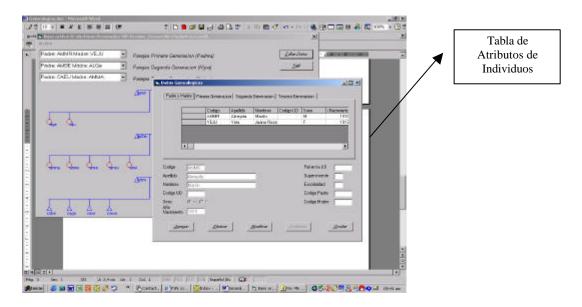


Fig. 2 La carga de los atributos de los individuos se desarrolla con el mismo nivel de detalle y con mayor comodidad que en la tabla maestra de la base de datos. Además de la simplicidad de ejecución de este proceso, lo que agrega esta forma de trabajo es la posibilidad de visualizar separadamente las generaciones de individuos de cada núcleo familiar. No es que solo se pueden cargar los datos de una manera más intuitiva. Lo importante es que la visualización de esta información se produce incorporando la semántica original con que los datos se someten a análisis. Cada generación se vincula con la anterior a través de un ancestro en común, lo que hace que la secuencia en que se ingresan los datos sea relacionalmente significativa.

La siguiente etapa de procesamiento de los datos permite generar lo que llamamos "matriz de parentesco extendido". Esta matriz se construye sumando todos los lazos parentales y asignando un valor a cada vínculo entre unidades familiares. Si ellas tienen al menos un miembro en común, están conectadas en la red, asumiendo un valor relacional de 1. En el caso contrario, de ausencia absoluta de conexión, el valor vinculante es 0. El programa construye estas redes automáticamente y retraduciendo el contenido de los datos básicos ingresados, sin ninguna intervención humana intermedia, pero lo que adicionalmente ofrece es la posibilidad de generar una matriz no dicotomizada, es decir una matriz que muestre exactamente la cantidad de miembros en común que tienen ambas unidades familiares.

La vinculación entre el esquema de trabajo original y el software se puede apreciar en la Figura 3:

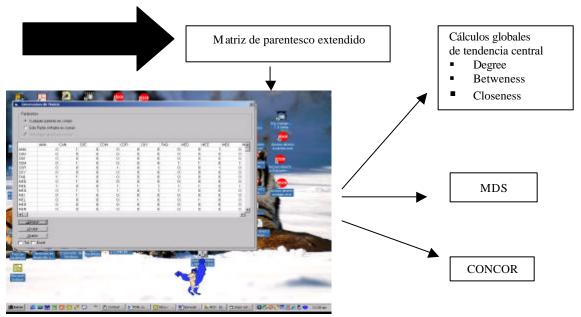


Fig. 3 Esta es la matriz de parentesco extendida en su versión dicotomizada, aunque también puede calcularse sin dicotomizar sus valores relacionales.

No todas las operaciones del esquema de trabajo original son soportadas por el programa. Como ya aclaramos más arriba, los cálculos correlacionales no son considerados y siguen siendo parte solo de nuestra investigación de origen, ya que responden a heurísticas de investigación más particulares.

Finalmente, el resultado de la representación analítica conseguida en el contexto de este software es compatible con programas de ARS, ya que las matrices pueden salvarse en formatos excel y texto. Sin mediación alguna, el UCINET V y el PAJEK pueden interpretar la información matricial y dar lugar a dispositivos propios de cálculo y tratamiento de la información generada. (Ver Fig. 4)

Representación Gráfica Paiek

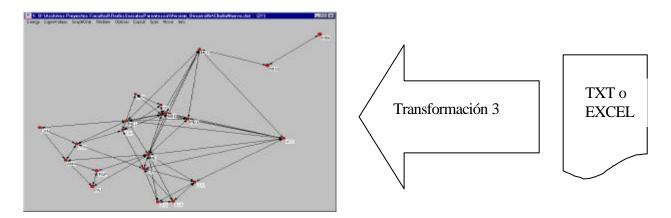


Fig. 4 Representación gráfica generada por PAJEK en base a un archivo de texto provisto por el programa de Genealogías

4- Conclusiones

El desarrollo de este software nos otorga varias ventajas sobre los métodos manuales de volcado de información. El proceso es mucho más veloz, consistente y confiable que apelando al ingreso manual, y además se pueden recorrer y modificar los datos de manera completamente intuitiva y clara, sin luchar con las dificultades inherentes a la interacción directa con la base de datos en su formato relacional. Además, la alimentación de información realizada por esta vía se desarrolla echando mano a la misma semántica del problema que se quiere analizar, cosa imposible de hacer si el usuario ejecuta la carga en la interfase del software final especializado en manejo de ARS.

En realidad nuestro programa posibilita tanto un acortamiento respecto de los tiempos de procesamiento de la información requeridos por un abordaje manual como un aumento en la confianza y consistencia de los datos obtenidos si se compara esta metodología con una aproximación tradicional.

Naturalmente, no siempre las mediciones o estadísticas realizadas iluminan aspectos desconocidos de los datos, y a veces simplemente sirven para detallar lo que la percepción intuitiva concibe de antemano, pero aunque

este sea el caso el servicio que esta clase de aplicaciones presta es enorme porque posibilita la exposición ordenada y consistente de cuestiones que de otro modo no podrían ser descriptas más que apelando al lenguaje natural y a partir de construcciones retóricas inmunes a cualquier práctica de contrastación. Los beneficios del uso de este tipo de recursos son ostensibles: se ahorra tiempo de trabajo en una medida sorprendente, se pueden construir y probar hipótesis con un grado de fiabilidad importante y se pueden replicar los procedimientos de refutación y corroboración de hipótesis tantas veces como se quiera, aunque esto no nos deje a salvo de las peripecias de cualquier proceso de investigación ligadas al valor de verdad de nuestras premisas.

En el caso puntual de nuestra investigación lo notable es que no hubo una única direccionalidad del trabajo de campo en relación al desarrollo analítico que propusimos después, y esto tal vez contribuya a cuestionar un poco la conocida suposición de que datos extraídos con un propósito determinado no puedan ser reutilizados con finalidades divergentes o encuadradas en otra concepción epistemológica. Ciertamente tampoco planteamos la prevalencia del abordaje de redes sociales con respecto a alguna otra estrategia de análisis estadístico no relacional, y mucho menos su incompatibilidad. La etapa de cálculos correspondiente a la teoría de redes sociales, por ejemplo, es muy relevante pero no es la única, ya que los cómputos correlacionales finales retoman los resultados de las medidas de tendencia central aplicada a los grupos pero vinculándolas a atributos que no son inferidos usando el Análisis de Redes Sociales. Ni siquiera estos métodos son incompatibles o muestran algún grado de inadecuación respecto de metodologías de recolección de datos clasificables como "cualitativas", ya que para el ARS representa exactamente lo mismo que la información haya sido obtenida permaneciendo en la comunidad o utilizando fuentes externas. Siempre que se disponga de un conjunto de datos traducibles a las entidades que definimos como nodos y a sus respectivos lazos, esta perspectiva será aplicable, y junto a ella toda la batería de herramientas de validación y taxonomización que hemos presentado.

Finalmente, quisiéramos remarcar la necesidad de evaluar las ventajas reales que ofrece la elaboración de un programa con prestaciones ad hoc.

En la temática que nos tocó abordar, los análisis de ARS sobre la población tehuelche del Chalía no demandaban necesariamente un entorno integrado de trabajo como el que propone el programa. El desarrollo del mismo responde más a necesidades de comodidad que a cuestiones de pertinencia analítica, como en casi todas las situaciones en las que un programa informático reemplaza total o parcialmente a la operatoria humana. Lo que simplifica el proceso de creación de un software, por más limitado que sea en sus prestaciones, es la existencia de un soporte previo que pertenezca al mismo contexto en que va a ser realizado. Nosotros ya disponíamos de información volcada en las estructuras tabulares de bases de datos que creímos adecuadas, y esto sirvió de gran ayuda en los pasos subsiguientes debido a que no tuvimos que partir de información no estructurada y heterogénea en sus fuentes.

En realidad el propio diseño de investigación, muy delimitado en sus alcances, fue el que hizo posible la implementación de un programa articulado funcionalmente a sus objetivos. Podríamos agregar, en este sentido, que los efectos epistemológicamente clarificadores que el desarrollo de un software provee son a veces más importantes que el ahorro de tiempo y el aumento de consistencia y falsabilidad que habilita.

Bibliografía

Godelier, Maurice (1971) "Modos de Producción relaciones de parentesco y estructuras demográficas" en Leach E. *El replanteamiento de la antropología* Barcelona. Seix Barral.

Lévi-Strauss, Claude (1969) *Las estructuras elementales del parentesco.* Buenos Aires. Paidos.

Muñiz, Marcelo (1998). Tesis de Licenciatura: Comunidad Tehuelche del Chalia: Aspectos económicos de su reproducción.

Muñiz, Marcelo. La economía de los pobladores del Chalía. (mimeo)

Redes - Revista Hispana para el análisis de redes sociales. Volumen 2#4

Rodriguez Diaz, Josep (1995). *Análisis estructural y de redes*, Madrid. Centro de Investigaciones Sociológicas.

Reynoso, Carlos (1986) *Teoría y Critica de la Antropología Cognitiva*, Buenos Aires. Ediciones Búsqueda.

Reynoso, Carlos (1998) *Corrientes en Antropología Contemporánea,* Buenos Aires. Editorial Biblos.

Quiroga A. Ortiz A. Miceli J. Muñiz M y Guerrero S. (1999) *Análisis de redes sociales en una comunidad tehuelche.* Ponencia presentada en la 3era reunión de antropología del Mercosur. Posadas.

Russell Bernard, H. (1995). *Research Methods in Anthropology*, Altamira Press.

Wasserman, Stanley and K. Faust (1994). *Social Network Analysis: Methods and Applications*, New York: Cambridge University Press.

Referencias a Internet

Sitio de Redes Sociales:

http://usuarios.lycos.es/redes/index.htm

Sitio para bajar programa de genealogías:

http://www11.brinkster.com/caos/index.asp